

English version of JP-B-2694618

W0433

R2!



Home

Search

List

 Include

## MicroPatent® PatSearch FullText: Record 1 of 1

Search scope: US Granted ; Full patent spec.

Years: 1971-2002

Text: Patent/Publication No.: 5095344

[no drawing available]

[Order This Patent](#)[Family Lookup](#)[Citation Indicators](#)[Go to first matching text](#)

US5095344 A1

Highly compact EPROM and flash EEPROM devices

Not Available

Inventor(s): Harari, Eliyahou

Application No. 204175, Filed 19880608, Issued 19920310

**Abstract:** Structures, methods of manufacturing and methods of use of electrically programmable read only memories (EPROM) and flash electrically erasable and programmable read only memories (EEPROM) include split channel and other cell configurations. An arrangement of elements and cooperative processes of manufacture provide self-alignment of the elements. An intelligent programming technique allows each memory cell to store more than the usual one bit of information. An intelligent erase algorithm prolongs the useful life of the memory cells. Use of these various features provides a memory having a very high storage density and a long life, making it particularly useful as a solid state memory in place of magnetic disk storage devices in computer systems.

**US.Class:** 257328 257327 257488 3651853 3651859 36518519 3651852 36518522 36518527  
3651853 36518531 36518533

**Int'l Class:** H01L02978; H01L02701 H01L02910 H01L02940

**Patents Cited:**

- EP0047153
- EP0219241
- JP58054668
- JP58121678
- JP62165370
- JP63093158
- US4331968
- US4361847
- US4377818
- US4412311
- US4422092
- US4462090
- US4486769
- US4503519
- US4577215
- US4665417

- US4717943
- US4763299
- US4794565
- US4803529
- US4852062
- US4935378

**Non-Patent Citations:**

- S. Tanaka et al., "A Programmable 256K CMOS EPROM with On-Chip Test Circuits", 1984 ISSCC Digest of Technical Papers, pp. 148-149.
- H. A. R. Wegener, "Endurance Model for Textured-Poly Floating Gate Memories", Technical Digest of the IEEE International Electron Device Meeting, Dec. 1984, pp. 480-483.
- Y. Mizutani and K. Makita, "A New EPROM Cell with a Side-Wall Floating Gate for High-Density and High-Performance Device", 1985 IEDM Technical Digest, pp. 635-638.
- F. Masuoka et al., "A 256K Flash EEPROM Using Triple Polysilicon Technology", Digest of Technical Papers, IEEE International Solid-State Circuits Conference, Feb. 1985, pp. 168-169, p. 335.
- A. T. Wu et al., "A Novel High-Speed, 5-Volt Programming EPROM Structure with Source-Side Injection", 1986 IEDM Technical Digest, pp. 584-587.
- G. Samachisa et al., "A 128K Flash EEPROM Using Double-Polysilicon Technology", IEEE Journal of Solid State Circuits, Oct. 1987, vol. SC-22, No. 5, pp. 676-683.
- H. Kume et al., "A Flash-Erase EEPROM Cell with an Asymmetrical Source and Drain Structure", Technical Digest of the IEEE International Electron Devices Meeting, Dec. 1987, pp. 560-563.
- V. N. Kynett et al., "An In-System Reprogrammable 256K CMOS Flash Memory", Digest of Technical Papers, IEEE International Solid-State Circuits Conference, Feb. 1988, pp. 132-133, 330.
- M. Horiguchi et al., "An Experimental Large-Capacity Semiconductor File Memory Using 16-Levels/Cell Storage", IEEE Journal of Solid-State Circuits, vol. 23, No. 1, Feb. 1988, pp. 27-33.
- T. Furuyama et al., "An Experimental 2-Bit/Cell Storage DRAM for Macro Cell or Memory-on-Logic Application", IEEE Custom Integrated Circuits Conference, May 1988, pp. 4.4.1-4.4.4.
- Muller et al., "Electrically Alterable 8192 Bit N-Channel MOS PROM", 1977 IEEE International Solid-State Circuits Conference, Feb. 18, 1977, pp. 188-189.

**Patents Citing This One (123):**

- US6243321B1 20010605 {n/a}  
Electrically alterable non-volatile memory with n-bits per cell
- US6272050B1 20010807 VLSI Technology, Inc.  
Method and apparatus for providing an embedded flash-EEPROM technology
- US6202138B1 20010313 Lexar Media, Inc  
Increasing the memory performance of flash memory devices by writing sectors simultaneously to multiple flash memory devices
- US5440505A1 19950808 Intel Corporation  
Method and circuitry for storing discrete amounts of charge in a single memory element
- US6343034B2 20020129 BTG International Inc.  
Electrically alterable non-volatile memory with n-bits per cell
- US5313427A1 19940517 Texas Instruments Incorporated  
EEPROM array with narrow margin of voltage thresholds after erase

- US6232633B1 20010515 International Business Machines Corporation  
NVRAM cell using sharp tip for tunnel erase
- US6151248A1 20001121 SanDisk Corporation  
Dual floating gate EEPROM cell array with steering gates shared by adjacent cells
- WO9810425A1 19980312 INTEL CORPORATION  
METHOD AND APPARATUS FOR CORRECTING A MULTILEVEL CELL MEMORY BY USING INTERLEAVING
- US6261903B1 20010717 Mosel Vitelic, Inc.  
Floating gate method and device
- US6091633A1 20000718 SanDisk Corporation  
Memory array architecture utilizing global bit lines shared by multiple cells
- US6246613B1 20010612 BTG International Inc.  
Memory apparatus including programmable non-volatile multi-bit memory cell, and apparatus and method for demarcating memory states of the cell
- US6136652A1 20001024 NotAvailable  
Preventing dielectric thickening over a channel area of a split-gate transistor
- US5862080A1 19990119 SanDisk Corporation  
Multi-state flash EEPROM system with defect handling
- US5910915A1 19990608 Sandisk Corporation  
EEPROM with split gate source side injection
- DE4407248A1 19950504 Gold Star Electron Co.  
{n/a}
- US6091618A1 20000718 Intel Corporation  
Method and circuitry for storing discrete amounts of charge in a single memory element
- US5476801A1 19951219 Cirrus Logic, Inc.  
Spacer flash cell process
- US6353554B1 20020305 BTG International Inc.  
Memory apparatus including programmable non-volatile multi-bit memory cell, and apparatus and method for demarcating memory states of the cell
- US5753525A1 19980519 International Business Machines Corporation  
Method of making EEPROM cell with improved coupling ratio
- US5554553A1 19960910 NotAvailable
- US5946234A1 19990831 Advanced Micro Devices, Inc.  
Constant current source programming of electrically programmable memory arrays
- US5687115A1 19971111 inVoice Technology, Inc.  
Write circuits for analog memory
- US5414829A1 19950509 Intel Corporation  
Override timing control circuitry and method for terminating program and erase sequences in a flash memory
- US6313498B1 20011106 Actrans System Inc.  
Flash memory cell with thin floating gate with rounded side wall, and fabrication process
- US5737265A1 19980407 Intel Corporation  
Programming flash memory using data stream analysis
- US5138576A1 19920811 Altera Corporation  
Method and apparatus for erasing an array of electrically erasable

EPROM cells  
→ US5602987A1 19970211 SanDisk Corporation  
Flash EEprom system  
→ US5455792A1 19951003 NotAvailable  
Flash EEPROM devices employing mid channel injection  
→ US6266278B1 20010724 SanDisk Corporation  
Dual floating gate EEPROM cell array with steering gates shared adjacent cells  
→ US6356486B1 20020312 BTG International Inc.  
Electrically alterable non-volatile memory with n-bits per cell  
→ US5936971A1 19990810 SanDisk Corporation  
Multi-state flash EEprom system with cache memory  
→ US5999446A1 19991207 SanDisk Corporation  
Multi-state flash EEprom system with selective multi-sector erase  
→ US5847996A1 19981208 Sandisk Corporation  
Eeprom with split gate source side injection  
→ US6104640A1 20000815 BTG International Inc.  
Electrically alterable non-violatile memory with N-bits per cell  
→ US5304505A1 19940419 NotAvailable  
Process for EEPROM cell structure and architecture with increased capacitance and with programming and erase terminals shared between several cells  
→ US5592002A1 19970107 NEC Corporation  
Non-volatile semiconductor memory device having reduced current consumption  
→ US5440518A1 19950808 NotAvailable  
Non-volatile memory circuits, architecture and methods  
→ US6222762B1 20010424 Sandisk Corporation  
Multi-state memory  
→ US6117733A1 20000912 Taiwan Semiconductor Manufacturing Company  
Poly tip formation and self-align source process for split-gate flash cell  
→ US6281075B1 20010828 SanDisk Corporation  
Method of controlling of floating gate oxide growth by use of an oxygen barrier  
→ US6317363B1 20011113 Sandisk Corporation  
Multi-state memory  
→ US6149316A1 20001121 SanDisk Corporation  
Flash EEprom system  
→ US6081447A1 20000627 Western Digital CorporationSanDisk Corporation  
Wear leveling techniques for flash EEPROM systems  
→ US5892710A1 19990406 Intel Corporation  
Method and circuitry for storing discrete amounts of charge in a single memory element  
→ US6266724B1 20010724 SanDisk Corporation  
Removable mother/daughter peripheral card  
→ US5781472A1 19980714 Intel Corporation  
Bit map addressing schemes for flash/memory  
→ US5268319A1 19931207 NotAvailable  
Highly compact EPROM and flash EEPROM devices  
→ US6275419B1 20010814 SanDisk Corporation  
Multi-state memory  
→ US5847425A1 19981208 SanDisk Corporation  
Dense vertical programmable read only memory cell structures and processes for making them

- US5648930A1 19970715 Symbios Logic Inc.  
Non-volatile memory which is programmable from a power source
- US5388083A1 19950207 Cirrus Logic, Inc.  
Flash memory mass storage architecture
- US5369615A1 19941129 SunDisk Corporation  
Method for optimum erasing of EEPROM
- US5828616A1 19981027 Intel Corporation  
Sensing scheme for flash memory with multilevel cells
- US6291297B1 20010918 Actrans System Inc.  
Flash memory cell with self-aligned gates and fabrication process
- US6339545B2 20020115 BTG International Inc.  
Electrically alterable non-volatile memory with n-bits per cell
- US5991517A1 19991123 SanDisk Corporation  
Flash EEPROM system with cell by cell programming verification
- US6327189B2 20011204 BTG International Inc.  
Electrically alterable non-volatile memory with n-bits per cell
- US6344993B1 20020205 SanDisk Corporation  
Dual floating gate EEPROM cell array with steering gates shared by adjacent cells
- US5756385A1 19980526 SanDisk Corporation  
Dense flash EEPROM cell array and peripheral supporting circuits formed in deposited field oxide with the use of spacers
- US5270979A1 19931214 SunDisk Corporation  
Method for optimum erasing of EEPROM
- US5867429A1 19990202 SanDisk Corporation  
High density non-volatile flash memory without adverse effects of electric field coupling between adjacent floating gates
- US6258668B1 20010710 Aplus Flash Technology, Inc.  
Array architecture and process flow of nonvolatile memory devices for mass storage applications
- US6329688B1 20011211 Mitsubishi Denki Kabushiki Kaisha  
Nonvolatile semiconductor memory device and method of manufacturing the same
- US5596486A1 19970121 Kaman Aerospace Corporation  
Hermetically sealed memory or PC card unit having a frame, header and covers in bonded engagement
- US5677869A1 19971014 Intel Corporation  
Programming flash memory using strict ordering of states
- US6230233B1 20010508 Sandisk Corporation  
Wear leveling techniques for flash EEPROM systems
- US5661053A1 19970826 SanDisk Corporation  
Method of making dense flash EEPROM cell array and peripheral supporting circuits formed in deposited field oxide with the use of spacers
- US5910912A1 19990608 International Business Machines Corporation  
Flash EEPROM with dual-sidewall gate
- US5701266A1 19971223 Intel Corporation  
Programming flash memory using distributed learning methods
- US6262452B1 20010717 Sanyo Electric Co., Ltd.  
Nonvolatile semiconductor memory device and manufacturing method therefor
- US5428621A1 19950627 SunDisk Corporation  
Latent defect handling in EEPROM devices
- US5386132A1 19950131 NotAvailable

- US6143609A1 Multimedia storage system with highly compact memory device  
20001107 Matsushita Electronics Corporation
- US6038166A1 Method for forming semiconductor memory device  
20000314 Invox Technology
- US6304485B1 High resolution multi-bit-per-cell memory  
20011016 SanDisk Corporation
- US5712179A1 Flash EEPROM system  
19980127 SanDisk Corporation
- US5534456A1 Method of making triple polysilicon flash EEPROM arrays having a separate erase gate for each row of floating gates  
19960709 SanDisk Corporation
- US6069039A1 Method of making dense flash EEPROM cell array and peripheral supporting circuits formed in deposited field oxide with sidewall spacers  
20000530 SanDisk Corporation
- US5776810A1 Plane decode/virtual sector architecture  
19980707 Sandisk Corporation
- US5267194A1 Method for forming EEPROM with split gate source side injection  
19931130 Winbond Electronics Corporation
- US5764571A1 Electrically erasable programmable read-only-memory cell with sidewall floating gate  
19980609 BTG USA Inc.
- US5659550A1 Electrically alterable non-volatile memory with N-bits per cell  
19970819 SanDisk Corporation
- US5519843A1 Latent defect handling in EEPROM devices  
19960521 M-Systems
- US6256755B1 Flash memory system providing both BIOS and user storage capability  
20010703 International Business Machines Corporation
- US6312989B1 Apparatus and method for detecting defective NVRAM cells  
20011106 Taiwan Semiconductor Manufacturing Company
- US5595924A1 Structure with protruding source in split-gate flash  
19970121 SanDisk Corporation
- US5479638A1 Technique of forming over an irregular surface a polysilicon layer with a smooth surface  
19951226 Cirrus Logic, Inc.
- US5729489A1 Flash memory mass storage architecture incorporation wear leveling technique  
19980317 Intel Corporation
- US6002152A1 Programming flash memory using predictive learning methods  
19991214 SanDisk Corporation
- US6002614A1 EEPROM with split gate source side injection with sidewall spacers  
19991214 BTG International Inc.
- US6201274B1 Memory apparatus including programmable non-volatile multi-bit memory cell, and apparatus and method for demarcating memory states of the cell  
20010313 NEC Corporation
- US6151254A1 Semiconductor device with no step between well regions  
20001121 NEC Corporation
- US6103573A1 Non-volatile semiconductor memory device and data erase method of non-volatile semiconductor memory device  
20000815 SanDisk Corporation
- US5361235A1 Processing techniques for making a dual floating gate EEPROM cell array  
19941101 NEC Corporation

Method for erasing data stored in a non-volatile semiconductor memory by using a predetermined series of pulses  
→ US6363008B1 20020326 Multi Level Memory Technology  
Multi-bit-cell non-volatile memory with maximized data capacity  
→ US5831901A1 19981103 Advanced Micro Devices, Inc.  
Method of programming a memory cell to contain multiple values  
→ US6028336A1 20000222 SanDisk Corporation  
Triple polysilicon flash EEPROM arrays having a separate erase gate for each row of floating gates, and methods of manufacturing such arrays  
→ US6259131B1 20010710 Taiwan Semiconductor Manufacturing Company  
Poly tip and self aligned source for split-gate flash cell  
→ US5815443A1 19980929 Intel Corporation  
Bit map addressing schemes for flash memory  
→ US5742543A1 19980421 Intel Corporation  
Flash memory device having a page mode of operation  
→ US5654217A1 19970805 SanDisk Corporation  
Dense flash EEPROM cell array and peripheral supporting circuits formed in deposited field oxide with the use of spacers  
→ US5566125A1 19961015 Intel Corporation  
Method and circuitry for storing discrete amounts of charge in a single memory element  
→ US5910925A1 19990608 Sandisk Corporation  
EEPROM with split gate source side injection  
→ US6317364B1 20011113 Sandisk Corporation  
Multi-state memory  
→ US5457606A1 19951010 Raymond Engineering Inc.  
Hermetically sealed PC card unit including a header secured to a connector  
→ US5268318A1 19931207 NotAvailable  
Highly compact EPROM and flash EEPROM devices  
→ US6040997A1 20000321 Lexar Media, Inc.  
Flash memory leveling architecture having no external latch  
→ US5838039A1 19981117 Matsushita Electronics Corporation  
Semiconductor memory having a tunneling region  
→ WO9419807A1 19940901 CONNER PERIPHERALS, INC.  
FLASH SOLID STATE DRIVE  
→ US5838616A1 19981117 Symbios, Inc.  
Gate edge aligned EEPROM transistor  
→ US5883409A1 19990316 Sandisk Corporation  
EEPROM with split gate source side injection  
→ US5539690A1 19960723 Intel Corporation  
Write verify schemes for flash memory with multilevel cells  
→ US5965913A1 19991012 SanDisk Corporation  
Dense vertical programmable read only memory cell structures and processes for making them  
→ US5754566A1 19980519 Intel Corporation  
Method and apparatus for correcting a multilevel cell memory by using interleaving  
→ US5748546A1 19980505 Intel Corporation  
Sensing scheme for flash memory with multilevel cells  
→ US5661687A1 19970826 Symbios Logic Inc.  
Drain excluded EEPROM cell  
→ US5747359A1 19980505 SanDisk Corporation

→ US6084797A1	Method of patterning polysilicon layers on substrate 20000704 STMicroelectronics S.r.l.
→ US6344998B2	Method for reading a multiple-level memory cell 20020205 BTG International Inc.
→ US5877986A1	Electrically alterable non-volatile memory with N-Bits per cell 19990302 SanDisk Corporation
	Multi-state Flash EEPROM system on a card that includes defective cell substitution
→ US5798968A1	19980825 SanDisk Corporation
	Plane decode/virtual sector architecture
→ US6324121B2	20011127 BTG International Inc.
	Electrically alterable non-volatile memory with n-bits per cell

Agent(s):Majestic, Parsons, Siebert & Hsue

Examiner(s):James, Andrew J.; Kim, Daniel

Go to Claims

## Background and Summary

### BACKGROUND OF THE INVENTION

This invention relates generally to semiconductor electrically programmable read only memories (EEPROM) and electrically erasable programmable read only memories (EEPROM), and specifically to semiconductor structures of such memories, processes of making them, and techniques for using them.

An electrically programmable read only memory (EPROM) utilizes a floating (unconnected) conductive gate, in a field effect transistor structure, positioned over but insulated from a channel region in a semiconductor substrate, between source and drain regions. A control gate is then provided over the floating gate, but also insulated therefrom. The threshold voltage characteristic of the transistor is controlled by the amount of charge that is retained on the floating gate. That is, the minimum amount of voltage (threshold) that must be applied to the control gate before the transistor is turned "on" to permit conduction between its source and drain regions is controlled by the level of charge on the floating gate. A transistor is programmed to one of two states by accelerating electrons from the substrate channel region, through a thin gate dielectric and onto the floating gate.

The memory cell transistor's state is read by placing an operating voltage across its source and drain and on its control gate, and then detecting the level of current flowing between the source and drain as to whether the device is programmed to be "on" or "off" at the control gate voltage selected. A specific, single cell in a two-dimensional array of EPROM cells is addressed for reading by application of a source-drain voltage to source and drain lines in a column containing the cell being addressed, and application of a control gate voltage to the control gates in a row containing the cell being addressed.

This type of EPROM transistor is usually implemented in one of two basic configurations. One is where the floating gate extends substantially entirely over the transistor's channel region between its source and drain. Another type, preferred in many applications, is where the floating gate extends from the drain region only part of the way across the channel. The control gate then extends completely across the channel, over the floating gate and then across the remaining portion of the channel not occupied by the floating gate. The control gate is separated

from that remaining channel portion by a thin gate oxide. This second type is termed a "split-channel" Eprom transistor. This results in a transistor structure that operates as two transistors in series, one having a varying threshold in response to the charge level on the floating gate, and another that is unaffected by the floating gate charge but rather which operates in response to the voltage on the control gate as in any normal field effect transistor.

Early Eprom devices were erasable by exposure to ultraviolet light. More recently, the transistor cells have been made to be electrically erasable, and thus termed electrically erasable and programmable read only memory (EEprom). One way in which the cell is erased electrically is by transfer of charge from the floating gate to the transistor drain through a very thin tunnel dielectric. This is accomplished by application of appropriate voltages to the transistor's source, drain and control gate. Other EEprom memory cells are provided with a separate, third gate for accomplishing the erasing. An erase gate passes through each memory cell transistor closely adjacent to a surface of the floating gate but insulated therefrom by a thin tunnel dielectric. Charge is then removed from the floating gate of a cell to the erase gate, when appropriate voltages are applied to all the transistor elements. An array of EEprom cells are generally referred to as a Flash EEprom array because an entire array of cells, or significant group of cells, is erased simultaneously (i.e., in a flash).

EEprom's have been found to have a limited effective life. The number of cycles of programming and erasing that such a device can endure before becoming degraded is finite. After a number of such cycles in excess of 10,000, depending upon its specific structure, its programmability can be reduced. Often, by the time the device has been put through such a cycle for over 100,000 times, it can no longer be programmed or erased properly. This is believed to be the result of electrons being trapped in the dielectric each time charge is transferred to or away from the floating gate by programming or erasing, respectively.

It is the primary object of the present invention to provide Eprom and EEprom cell and array structures and processes for making them that result in cells of reduced size so their density on a semiconductor chip can be increased. It is also an object of the invention that the structures be highly manufacturable, reliable, scalable, repeatable and producible with a very high yield.

It is yet another object of the present invention to provide EEprom semiconductor chips that are useful for solid state memory to replace magnetic disk storage devices.

Another object of the present invention is to provide a technique for increasing the amount of information that can be stored in a given size Eprom or EEprom array.

Further, it is an object of the present invention to provide a technique for increasing the number of program/read cycles that an EEprom can endure.

## SUMMARY OF THE INVENTION

These and additional objects are accomplished by the various aspects of the present invention, either alone or in combination, the primary aspects being briefly summarized as below:

1. The problems associated with prior art split channel Eprom and split channel Flash EEprom devices are overcome by providing a split channel memory cell constructed in one of the following ways:

(A) In one embodiment, one edge of the floating gate is self aligned to and overlaps the edge of the drain diffusion and the second edge of the floating gate is self aligned to but is spaced apart from the edge of the source diffusion. A sidewall spacer formed along the second edge of the

floating gate facing the source side is used to define the degree of spacing between the two edges. Self alignment of both source and drain to the edges of the floating gate results in a split channel Eprom device having accurate control of the three most critical device parameters: Channel segment lengths L1 and L2 controllable by floating gate and control gate, respectively, and the extent of overlap between the floating gate and the drain diffusion. All three parameters are insensitive to mask misalignment and can be made reproducibly very small in scaled-down devices.

(B) In a second embodiment of the split channel Eprom a heavily doped portion of the channel adjacent to the drain diffusion is formed by a novel, well-controlled technique. The length  $L_p$  and doping concentration of this channel portion become the dominant parameters for programming and reading, thereby permitting the formation of a split channel structure which is relatively insensitive to misalignments between the floating gate and the source/drain regions.

2. A separate erase gate is provided to transform a Eprom device into a Flash EEprom device. The area of overlap between the floating gate and the erase gate is insensitive to mask misalignment and can therefore be made reproducibly very small.

3. In some embodiments of this invention, the erase gate is also used as a field plate to provide very compact electric isolation between adjacent cells in a memory array.

4. A new erase mechanism is provided which employs tailoring of the edges of a very thin floating gate so as to enhance their effectiveness as electron injectors.

5. A novel intelligent programming and sensing technique is provided which permits the practical implementation of multiple state storage wherein each Eprom or flash EEprom cell stores more than one bit per cell.

6. A novel intelligent erase algorithm is provided which results in a significant reduction in the electrical stress experienced by the erase tunnel dielectric and results in much higher endurance to program/erase cycling.

The combination of various of these features results in new split channel Eprom or split channel Flash EEprom devices which are highly manufacturable, highly scalable, and offering greater storage density as well as greater reliability than any prior art Eprom or Flash EEprom devices. Memories that utilize the various aspects of this invention are especially useful in computer systems to replace existing magnetic storage media (hard disks and floppy disks), primarily because of the very high density of information that may be stored in them.

Additional objects, features and advantages of the present invention will be understood from the following description of its preferred embodiments, which description should be taken in conjunction with the accompanying drawings.

## Detailed Description

### DETAILED DESCRIPTION OF THE PRIOR ART

There are two distinctly different approaches in the prior art of Flash EEproms. A triple polysilicon device was described by J. Kupec et al. in 1980 IEDM Technical Digest, p. 602 in an article entitled "Triple Level Polysilicon EEprom with Single Transistor per Bit". An improvement to the Kupec device was proposed by F. Masuoka and H. Iizuka in U.S. Pat. No. 4,531,203, issued July 23, 1985. Variations on the same cell are described by C. K. Kuo and S. C. Tsaur in U.S. Pat. No. 4,561,004 issued Dec. 24, 1985, and by F. Masuoka et al. in an article titled "A 256K Flash

EEprom Using Triple Polysilicon Technology", Digest of Technical Papers, IEEE International Solid-State Circuits Conference, February 1985, p. 168.

The second approach is a double polysilicon cell described by G. Samachisa et al., in an article titled "A 128K Flash EEprom Using Double Polysilicon Technology", IEEE Journal of Solid State Circuits, October 1987, Vol. SC-22, No. 5, p. 676. Variations on this second cell are also described by H. Kume et al. in an article titled "A Flash-Erase EEprom Cell with an Asymmetric Source and Drain Structure", Technical Digest of the IEEE International Electron Devices Meeting, December 1987, p. 560, and by V. N. Kynett et al. in an article titled "An In-System Reprogrammable 256K CMOS Flash Memory", Digest of Technical Papers, IEEE International Solid-State Circuits Conference, February 1988, p. 132. A cross-section of the Samachisa cell is shown in FIG. 1. Transistor 100 is an NMOS transistor with source 101, drain 102, substrate 103, floating gate 104 and control gate 109. The transistor has a split channel consisting of a section 112 (L1) whose conductivity is controlled by floating gate 104, in series with a section 120 (L2) whose conductivity is controlled by control gate 109. Programming takes place as in other Eeprom cells by injection of hot electrons 107 from the channel at the pinchoff region 119 near the drain junction. Injected electrons are trapped on floating gate 104 and raise the conduction threshold voltage of channel region 112 and therefore of transistor 100. To erase transistor 100 the oxide in region 112 separating between the floating gate 104 and drain diffusion 102 and channel 112 is thinned to between 15 and 20 nanometers, to allow electronic tunneling of trapped electrons 108 from the floating gate to the drain. In the Samachisa cell the appropriate voltages applied to achieve programming are  $V_{sub.CG} = 12V$ ,  $V_{sub.D} = 9V$ ,  $V_{sub.BB} = 0V$ ,  $V_{sub.S} = 0V$ , and to achieve erase are  $V_{sub.CG} = 0V$ ,  $V_{sub.D} = 19V$ ,  $V_{sub.BB} = 0V$ ,  $V_{sub.S} = \text{floating}$ . Samachisa points out that the electrical erase is not self-limiting. It is possible to overerase the cell, leaving the floating gate positively charged, thus turning the channel portion L1 into a depletion mode transistor. The series enhancement transistor L2 is needed therefore to prevent transistor leakage in the overerase condition.

The Samachisa cell suffers from certain disadvantages. These are:

- (a) It is difficult to prevent avalanche junction breakdown or high junction leakage current at the drain junction 102 during the time the very high erase voltage is applied to the drain;
- (b) It is difficult to grow with high yields the thin oxide layer 112 used for tunnel erase;
- (c) Because of the presence of thin oxide layer between the floating gate and the drain diffusion, it is difficult to prevent accidental tunneling of electrons from the floating gate to the drain in what is known as the "program disturb" condition. Under this condition an unselected cell in a memory array sharing the same drain (bit line) as a programmed cell may have a drain voltage of approximately 10 volts and a control gate voltage of 0 volts. Although this represents a much weaker electric field than that experienced during tunnel erase (when the drain is at approximately 19 volts), it nevertheless can, over a prolonged period of time alter by slow tunneling the charge stored on the floating gate.

The Kynett and Kume cells (FIG. 2a) are similar to the Samachisa cell except for the elimination of the series enhancement transistor 120, and the performing of tunnel erase 208 over the source diffusion 201 rather than over the drain diffusion 202. Typically the Kynett cell uses during programming voltages  $V_{sub.CG} = 12V$ ,  $V_{sub.D} = 8V$ ,  $V_{sub.S} = 0V$ ,  $V_{sub.BB} = 0V$ , and during erase voltages  $V_{sub.S} = 12V$ ,  $V_{sub.BB} = 0V$ ,  $V_{sub.CG} = 0V$ ,  $V_{sub.D} = \text{floating}$ . Kynett achieves a lower erase voltage than Samachisa by thinning tunnel dielectric 212 to 10 nanometers or less, so that even though the voltage applied to the source diffusion during erase is reduced, the electric field across tunnel dielectric 212 remains as high as in the case of the Samachisa cell.

The Kynett cell can be contrasted with the Samachisa cell:

- (a) Kynett is less susceptible to avalanche breakdown of source diffusion 201 during erase because the voltage is reduced from 19 volts to 12 volts.
- (b) Kynett's cell is more susceptible to low yields due to pinholes in the thin dielectric layer 212 because its thickness is reduced from approximately 20 nanometers to approximately 10 nanometers.
- (c) Because Kynett uses a lower voltage for erase but essentially the same drain voltage for programming Kynett is far more susceptible to accidental "program disturb" due to partial tunnel erase (during programming) occurring from floating gate 204 to drain 202.
- (d) Kynett's cell is highly susceptible to an overerase condition because it does not have the series enhancement channel portion 120 of Samachisa's cell. To prevent overerase Kynett et al. deploy a special erase algorithm. This algorithm applies a short erase pulse to an array of cells, then measures the threshold voltage of all cells to ensure that no cell has been overerased into depletion. It then applies a second erase pulse and repeats the reading of all cells in the array. This cycle is stopped as soon as the last cell in the array has been erased to a reference enhancement voltage threshold level. The problem with this approach is that the first cell to have been adequately erased continues to receive erase pulses until the last cell has been adequately erased, and may therefore be susceptible to overerase into a depletion threshold state.

Kupec's cell employs essentially the Kynett cell without a thin tunnel dielectric over the source, channel, or drain, and with a third polysilicon plate covering the entire transistor and acting as an erase plate. A cross sectional view of the Kupec device is shown in FIG. 2b. Transistor 200b consists of a stacked floating gate 204b and control gate 209b with source 201b and drain 202b self aligned to the edges of the floating gate. Gate dielectric 212 is relatively thick and does not permit tunnel erase from floating gate to source or drain. An erase plate 230b overlies the control gate and covers the sidewalls of both the control gate and the floating gate. Erase takes place by tunneling across the relatively thick oxide 231b between the edges of floating gate 204b and erase plate 230b. Kupec attempts to overcome the overerase condition by connecting the erase plate during high voltage erase to drain 202b and through a high impedance resistor R (FIG. 2c) to the erase supply voltage V.sub.ERASE. As soon as the cell is erased into depletion the drain to source transistor conduction current drops most of the erase voltage across the resistor, reducing the voltage on the erase plate 230b to below the tunneling voltage. This approach is extremely difficult to implement in a block erase of a large array because different transistors begin conduction at different times.

Masuoka's approach to Flash EEPROM overcomes most of the disadvantages of the Samachisa, Kynett and Kupec cells. FIG. 3a provides a top view of the Masuoka prior art cell, FIG. 3b shows the schematic representation of the same cell, and FIG. 3c provides a cross section view along the channel from source to drain. Transistor 300 consists of a split channel EEPROM transistor having a source 301, a drain 302, a floating gate 304 controlling channel conduction along section L1 (312) of the channel, a control gate 309 capacitively coupled to the floating gate and also controlling the conduction along the series portion of the channel L2 (320), which has enhancement threshold voltage.

The transistor channel width (W), as well as the edges of the source and drain diffusions are defined by the edges 305 of a thick field oxide formed by isoplanar oxidation. Oxide 332 of thickness in the 25 to 40 nanometers range is used as isolation between the floating gate and the substrate. Masuoka adds an erase gate 330 disposed underneath the floating gate along one

of its edges. This erase gate is used to electrically erase floating gate 304 in an area of tunnel dielectric 331 where the floating gate overlaps the erase gate. Tunnel dielectric 331 is of thickness between 30 and 60 nanometers.

Masuoka specifies the following voltages during erase: V<sub>sub.S</sub> = 0V, V<sub>sub.D</sub> = 0V, V<sub>sub.CG</sub> = 0V, V<sub>sub.BB</sub> = 0V, V<sub>sub.ERASE</sub> = 20V to 30V.

Comparing the Masuoka cell with the Samachisa and Kynett cells:

- (a) Masuoka's cell does not erase by using either the source diffusion or the drain diffusion for tunnel erase. Therefore these diffusions never experience a voltage higher than during Eprom programming. The junction avalanche breakdown and junction leakage problems therefore do not exist.
- (b) Masuoka's cell uses a relatively thick tunnel dielectric and therefore does not need to use thin tunnel dielectrics for erase. Therefore it is less susceptible to oxide pinholes introduced during the manufacturing cycle.
- (c) Masuoka's cell does not have a "program disturb" problem because programming and tunnel erase involve two different mechanisms occurring at two different regions of the transistor.
- (d) Masuoka's cell is not susceptible to the overerase condition because of the presence of the series enhancement transistor channel 320 (L2).
- (e) Masuoka's cell requires a third layer of polysilicon, which complicates the process as well as aggravates the surface topology. Because the erase gate consumes surface area over the field oxide 305 it results in a larger cell.
- (f) The overlap area 331 in Masuoka's cell is sensitive to mask misalignment between the two masks defining this overlap. Since the overlap area is nominally very small, even small misalignments can result in large variations in the area used for tunnel erase. This results in severe variations from wafer to wafer.

From the foregoing analysis it is clear that while the Masuoka prior art cell successfully addresses most of the problems encountered by Samachisa and Kynett, it itself has disadvantages not encountered by Samachisa or Kynett.

Masuoka and Samaciisa both use a split channel Eprom transistor for programming. In the split channel eprom transistor, the portion L2 of the channel length controlled by control gate 109, 309 has a fixed enhancement threshold voltage determined by the p+ channel doping concentration 360. The portion L1 of the channel length controlled by floating gate 104 (Samachisa) and 304 (Masuoka) has a variable threshold voltage determined by the net charge stored on the floating gate.

Other prior art split channel Eprom transistors are described by E. Harari in U.S. Pat. No. 4,328,565 May 4, 1982 and by B. Eitan in U.S. Pat. No. 4,639,893, Jan. 27, 1987. The Harari split channel Eprom transistor 300d is shown in cross section in FIG. 3d. Source 301d and drain 302d are formed prior to formation of the floating gate 304d. Therefore, the total channel length L1+L2 is insensitive to mask misalignment. However, both L1 and L2 are sensitive to misalignment between floating gate 304d and drain diffusion 302d.

The Eitan split channel Eprom transistor 400 is shown in cross sections in FIG. 4a. The Eitan patent highlights the main reasons for using a split channel architecture rather than the standard

self aligned stacked gate Eprom transistor 200 (FIG. 2). These reasons can be summarized as follows:

The addition of a fixed threshold enhancement transistor in series with the floating gate transistor decouples the floating gate from the source diffusion. This allows the channel length L1 to be made very small without encountering punchthrough between source and drain. Furthermore, transistor drain-turnon due to the parasitic capacitive coupling between the drain diffusion and the floating gate is eliminated because the enhancement channel portion L2 remains off.

Eitan shows that the shorter the length L1 the greater the programming efficiency and the greater the read current of the split channel Eprom transistor. For Flash EEPROM devices the series enhancement channel L2 acquires additional importance because it allows the floating gate portion L1 to be overerased into depletion threshold voltage without turning on the composite split channel transistor.

The disadvantages incurred by the addition of the series enhancement channel L2 are an increase in cell area, a decrease in transistor transconductance, an increase in control gate capacitance, and an increase in variability of device characteristics for programming and reading brought about by the fact that L1 or L2 or both are not precisely controlled in the manufacturing process of the prior art split channel devices. Samachisa, Masuoka and Eitan each adopt a different approach to reduce the variability of L1 and L2:

Samachisa's transistor 100 (FIG. 1) uses the two edges 140, 143 of control gate 100 to define (by a self aligned ion implant) drain diffusion 102 and source diffusion 101. Edge 141 of floating gate 104 is etched prior to ion implant, using edge 140 of control gate 109 as an etch mask. This results in a split channel transistor where (L1+L2) is accurately controlled by the length between the two edges 140, 143 of the control gate. However, L1 and L2 are both sensitive to misalignment between the mask defining edge 142 and the mask defining edges 140, 143.

Masuoka's transistor 300 (FIG. 3c) forms both edges 341, 342 of floating gate 304 in a single masking step. Therefore L1 is insensitive to mask misalignment. L2, which is formed by ion implant of source diffusion 301 to be self aligned to edge 343 of control gate 309, is sensitive to misalignment between the mask defining edge 342 and the mask defining edge 343. Furthermore the Masuoka transistor 300 may form a third channel region, L3, if edge 340 of control gate 309 is misaligned in a direction away from edge 341 of floating gate 304, the formation of L3 will severely degrade the programming efficiency of such a cell.

Eitan's transistor 400 (FIGS. 4a, 4b) uses a separate mask layer 480 to expose the edge of floating gate 404 to allow drain diffusion 402 to be self aligned (by ion implantation) to edge 441 of floating gate 404. Therefore L1 can be accurately controlled and is not sensitive to mask misalignment. L2 however is sensitive to the misalignment between edge 482 of photoresist 480 and edge 442 of the floating gate. Eitan claims that the variability in L2 due to this mask misalignment, can be as much as 1.0 micron or more without affecting the performance of the device (see claims 3,4 of the above-referenced Eitan patent).

It should be pointed out that even with the most advanced optical lithography systems available today in a production environment it is difficult to achieve an alignment accuracy of better than +0.25 microns between any two mask layers. Therefore the variability in L2 or L1 inherent to any structure which is alignment sensitive can be as much as approximately 0.5 microns from one extreme to the other.

Another prior art split channel Eprom device which attempts to achieve the objective of

accurately establishing L1 and L2 is disclosed by Y. Mizutani and K. Makita in the 1985 IEDM Technical Digest, pp. 635- 638, shown in cross section in FIG. 4c. Transistor 400c has a floating gate 404c formed along the sidewall 440c of control gate 409c. In this way both L1 and L2 can be independently established and are not sensitive to mask misalignment. Transistor 400c has the drawback that the capacitive coupling between control gate 409c and floating gate 404c is limited to the capacitor area of the sidewall shared between them, which is relatively a small area. Therefore there is a very weak capacitive coupling between the control gate and the floating gate either during programming or during read. Therefore, although the device achieves good control of L1 and L2 it is of rather low efficiency for both modes of operation.

Yet another prior art device which has a split channel with a well controlled L1 and L2 is disclosed by A. T. Wu et al. in the 1986 IEDM Technical Digest, p. 584 in an article entitled "A Novel High-Speed, 5-Volt Programming Eprom Structure with Source-Side Injection". A cross section of the Wu prior art transistor is shown in FIG. 4d (FIG. 2 in the above-referenced article). This transistor has a floating gate 404d coupled to a control gate 409d, extending over channel region L1 (412d), in series with a second floating gate 492d formed in a sidewall adjacent to source diffusion 401d and overlying channel region L2 (420d). This second floating gate is capacitively coupled to the control gate 409d through the relatively small area of the sidewall 493d shared between them and is therefore only marginally better than the Mizutani prior art device, although it does achieve a good control of both L1 and L2.

Another prior art Eprom transistor which does not have a split channel structure but which seeks to achieve two distinct channel regions to optimize the Eprom programming performance is disclosed by S. Tanaka et al. in 1984 ISSCC Digest of Technical Papers, p. 148 in an article entitled "A Programmable 256K CMOS Eprom with On Chip Test Circuits". A cross section of this device is shown in FIG. 4e (corresponding to FIG. 3 in the Tanaka article). Transistor 400e is a stacked gate Eprom transistor (not split channel) with source 401e and drain 402e self aligned to both edges of floating gate 404e and control gate 409e. The channel region is more heavily p doped 460e than the p substrate 463e, but in addition there is a second p+ region 477e which is even more heavily p-doped than region 460e. This region 477e is formed by diffusion of boron down and sideways from the top surface on the drain side only, and is formed after formation of the floating gate so as to be self aligned to the floating gate on the drain side. The extent of sideway diffusion of boron ahead of the sideway diffusion of arsenic, from the N+ drain junction defines a channel region Lp (478e) adjacent to the drain. This is a DMOS type structure, called DSA (Diffusion Self Aligned) by Tanaka. The presence of the p+ region 478e reduces considerably the width of the drain depletion region during high voltage programming. A shorter depletion layer width results in greater energy being imparted to channel electrons entering the depletion region, which in turn results in significant increase in programming efficiency through hot electron injection. Transistor 400e has proven difficult to manufacture because it is rather difficult to control the length Lp and the surface channel concentration p+ through a double diffusion step. Furthermore, it is rather difficult to obtain value of Lp bigger than approximately 0.3 microns by diffusion because device scaling dictates the use of rather low temperature diffusion cycles. Still further, the DSA Eprom device suffers from an excessively high transistor threshold voltage in the unprogrammed (conducting) state, as well as from high drain junction capacitance. Both these effects can increase substantially the read access time.

#### Detailed Description of Specific Embodiments of the Invention

##### I.a. Split Channel Eprom Transistor with Self Aligned Drain Diffusion and Self Aligned Spaced Apart Source Diffusion

FIG. 5a presents a cross sectional view of a split channel Eprom transistor in accordance with a first embodiment of this invention. Transistor 500a consists of a p type silicon substrate 563 (which can alternatively be a p type epitaxial layer grown on top of a p++ doped silicon

substrate), N+ source diffusion 501a, N+ drain diffusion 502a, a channel region 560a which is more heavily p-doped than the surrounding substrate, a floating gate 504a overlying a portion L1 of the channel, 512a, and a control gate 509 overlying the remaining portion L2 of the channel, 520a as well as the floating gate. Floating gate 504a is dielectrically isolated from the surface of the silicon substrate by dielectric film 564a, which is thermally grown Silicon Dioxide. Control gate 509 is capacitively coupled to floating gate 504a through dielectric film 567a, which can either be thermally grown Silicon Dioxide or a combination of thin layers of Silicon Dioxide and Silicon Nitride. Control gate 509 is also insulated from the silicon surface in channel portion L2 as well as over the source and drain diffusions by dielectric film 565a, which is made of the same material as dielectric 567a.

P-type substrate 563 is typically 5 to 50 Ohms centimeter, p+ channel doping 560a is typically in the range of  $1 \times 10^{16}$  cm.<sup>-3</sup> to  $2 \times 10^{17}$  cm.<sup>-3</sup>, dielectric film 564a is typically 20 to 40 nanometers thick, dielectric film 567a is typically 20 to 50 nanometers thick, floating gate 504a is usually a heavily N+ doped film of polysilicon of thickness which can be as low as 25 nanometers (this thickness will be discussed in Section VII) or as high as 400 nanometers. Control gate 509 is either a heavily N+ doped film of polysilicon or a low resistivity interconnect material such as a silicide or a refractory metal. Of importance, edge 523a of N+ drain diffusion 502a formed by ion implantation of Arsenic or Phosphorus is self aligned to edge 522a of floating gate 504a, while edge 521a of N+ source diffusion 501a formed by the same ion implantation step is self aligned to, but is spaced apart from, edge 550a of the same floating gate 504a, using a sidewall spacer (not shown in FIG. 5a) which is removed after the ion implantation but prior to formation of control gate 509. The implant dose used to form diffusions 501a, 502a, is typically in the range of  $1 \times 10^{15}$  cm.<sup>-2</sup> to  $1 \times 10^{16}$  cm.<sup>-2</sup>.

The key steps for the formation of channel portions L1 and L2 are illustrated in FIGS. 5b through 5f. In the structure of FIG. 5b floating gates 504a, 504b are formed in a layer of N+ doped polysilicon on top of a thin gate oxide 564a, by anisotropic reactive ion etchings, using photoresist layer 590 as a mask. In FIG. 5c a thin protective film 566a is deposited or thermally grown, followed by the deposition of a thick spacer layer 570. The purpose of film 566a is to protect the underlying structure such as layer 565a from being etched or attacked when the spacer film is etched back. The spacer film is now etched back in an anisotropic reaction ion etch step with carefully controlled timing. The conditions for etchback must have no significant undercutting and must have a differential etch rate of 20:1 or higher between the spacer material and the material of protective film 566a. Spacer layer 570 can be a conformal film of undoped LPCVD polysilicon while protective film 566a can be silicon dioxide or silicon nitride. Alternatively, spacer layer 570 can be a conformal film of LPCVD silicon dioxide while protective film 566a can be either LPCVD silicon nitride or LPCVD polysilicon. The thickness of protective film 566a should be as thin as possible, typically in the range of 10 to 30 nanometers, so as to allow penetration of the subsequent Arsenic implantation to form the source and drain diffusions.

The thickness of the conformal spacer layer determines the width of the sidewall spacer, and therefore also the length of channel portion L2. Typically for an L2 of 400 nanometers a spacer layer of approximately 600 nanometers thickness is used.

In FIG. 5d spacers 592a, 593a and 592b, 593b are formed along the vertical edges of floating gates 504a and 504b respectively at the completion of the timed reactive ion etch step. These spacers result from the fact that the thickness of layer 570 is greater adjacent to the vertical walls of the floating gates than it is on flat surfaces. Therefore a carefully timed anisotropic reactive ion etchback will etch through layer 570 in areas of flat surface topology while not completely etching through it along each edge, forming the spacers. The technique for formation of narrow sidewall spacers along both edges of the gate of MOS transistors is well known in the industry, and is commonly used to form lightly doped drains (LDD) in short channel MOSFETS. (See, for example, FIG. 1 in an article in 1984 IEDM Technical Digest, p. 59 by S. Meguro et al.

titled "Hi-CMOS III Technology".)

In the present invention, the spacer can be significantly wider, it is used along one edge only, and it is used not to define a lightly doped source or drain but rather to define the series enhancement transistor channel portion L2.

The next step is a masking step. Photoresist 591a, 591b (FIG. 5d) is used as a mask to protect spacers 592a, 592b while exposing spacers 593a, 593b. The latter are etched away, preferably with a wet chemical etch (which should be chosen so as to not etch protective film 566a), and the photoresist is stripped.

In FIG. 5e ion implantation of Arsenic through dielectric films 566a and 565a is used to form N+ source diffusions 501a, 501b and N+ drain diffusions 502a, 502b. On the drain side these diffusions are self aligned to edges 522a and 522b of the floating gates. On the source side the diffusions are self aligned to edges 550a and 550b of the floating gates but are spaced apart from these edges by the width of spacers 592a and 592b less the sideways diffusion in subsequent high temperature process steps.

Next, spacers 592a, 592b and the protective film 566a are removed (FIG. 5f), preferably with wet etches which will not attack the underlying layers 565a and 504a. Dielectric film 567a is grown by thermal oxidation or deposited by LPCVD on the exposed surfaces of the floating gates and substrate. A conductive layer is then deposited and control gates 509a, 509b are formed through etching of long narrow strips which constitute the word lines in rows of memory cells in an array.

The remaining part of the process is standard:

The surface of the structure is covered with a thick passivation layer 568, usually phosphorous doped glass or a Borophosphosilicate glass (BPSG). This passivation is made to flow in a high temperature anneal step. Contact vias are etched (not shown in FIG. 5f) to allow electrical access to the source and drain diffusions. Metallic interconnect strips 569a, 569b are provided on top of passivation layer 568, accessing the source and drain diffusions through the via openings (not shown).

Comparing split channel transistor 500a of FIG. 5f with the Samachisa, Masuoka, Harari and Eitan prior art split channel transistors 100, 300, 300d and 400, the advantages of transistor 500a can be summarized as follows:

- a) L1 and L2 are insensitive to mark misalignment. Therefore they can be controlled much more accurately and reproducibly than the prior art.
- b) Because all four prior art transistors 100, 300, 300d and 400 define L2 through a mask alignment tolerance whereas transistor 500a defines L2 through control of the width of a sidewall spacer it is possible in transistor 500a to achieve controllably a much shorter channel portion L2 than possible through a mask alignment. This becomes an important consideration in highly scaled split channel Eprom and Flash EEprom transistors.

#### I.b. Split Channel Eprom Transistor with Heavily Doped Channel Adjacent to the Drain Junction

FIG. 14c presents a cross sectional view of a non self aligned split channel Eprom transistor in accordance with a second embodiment of this invention. FIGS. 14a and 14b illustrate the critical process steps in the manufacturing process of this device. Transistor 1400 consists of a p type silicon substrate 1463 (which can also be a p type epitaxial layer grown on a p++ substrate). Shallow N+ source diffusions 1401 and N+ drain diffusions 1402 are formed prior to formation of

floating gate 1404, in contrast with the embodiment of section Ia above. The channel region between the source and drain diffusions is split into two portions: a portion L1 (1412) which is lying directly underneath the floating gate, and a portion L2 (1420) which is lying directly underneath the control gate 1409. The improvement over the Harari prior art split channel transistor 300d (FIG. 3d) consists of a heavily p+ doped narrow region 1460 adjacent to drain diffusion 1402. The width Lp (1413) and doping concentration of this region at the top surface where the field effect transistor channel is formed, become the controlling parameters for device programming and reading efficiency, provided that p+ is sufficiently high. Typically, p substrate 1463 may have a p type doping concentration of 1.times.10.sup.16 cm.sup.-3 whereas p+ region 1460 may have a p+ type doping concentration of between 1.times.10.sup.17 cm.sup.-3 and 1.times.10.sup.18 cm.sup.-3. In the preferred manufacturing process the length Lp and doping concentration of region 1460 are chosen so that the depletion region width at the drain junction under programming voltage conditions is less than the width Lp. So long as that condition is satisfied, and so long as L1 is bigger than Lp, then the actual value of L1 is of secondary importance to the device performance. Since L1 in this device is determined through a mask alignment between the floating gate and the drain it is not as well controlled as in the Eitan prior art transistor 400. However, to the extent that region 1460 can be made to be self aligned to the drain so that parameter Lp is not sensitive to mask alignment, then any variability in L1 is of secondary importance, Lp being the controlling parameter.

A new method is disclosed for manufacturing the split channel Eprom transistor 1400 which results in much better control of the parameter Lp and of the surface channel doping concentration 1413 than is provided by the DSA (Diffusion Self Align) approach of the Tanaka prior art transistor 400e (FIG. 4e).

The main steps in this new method for the fabrication of a memory array of transistors 1400 are as follows:

1. In the structure of FIG. 14a a thin oxide layer 1475, typically 50 nanometers of silicon dioxide, is covered with a layer 1474 of silicon nitride, approximately 100 nanometers thick. This in turn is covered with a second layer 1473 of deposited silicon dioxide, approximately 100 nanometers thick. Oxide 1475 and nitride 1474 can, for example, be the same films used to form isoplanar isolation regions in the periphery of the memory array.
2. A photoresist mask P.R.1 (1470) is used to define source and drain regions in long parallel strips extending in width between edges 1471, 1472 of openings in the photoresist. Exposed oxide layer 1473 is now wet etched in a carefully controlled and timed etch step which includes substantial undercutting of photoresist 1470. The extent of undercutting, which is measured by the distance Lx between oxide edges 1476 and 1478, will eventually determine the magnitude of parameter Lp. Typically, Lx is chosen between 300 nanometers and 700 nanometers. The three parameters critical for a reproducible Lx are the concentration and temperature of the etch solution (hydrofluoric acid) and the density (i.e., lack of porosity) of the oxide 1473 being etched. These can be well controlled sufficiently so that a timed undercutting etch step results in well controlled etched strips of width Lx and running parallel to edges 1471, 1472 of the long openings in the photoresist. In fact, for values of Lx below approximately 500 nanometers, it is easier to achieve a reproducible Lx through controlled sideway etching than by controlling the line width of long, narrow line in a photoresist layer. An example of the use of sideway etching self aligned to an edge in a similar fashion (but to achieve the different purpose of forming a very narrow guard ring) can be found in the prior art article by S. Kim titled "A Very Small Schotky Barrier Diode with Self-Aligned Guard Ring for VLSI Application", appearing in the 1979 IEDM Technical Digest, p. 49.
3. At the completion of the sideway etch step a second, anisotropic etch is performed, using the same photoresist mask P.R.1 to etch away long strips of the exposed silicon nitride film 1474.

Edges 1471, 1472 of P.R.1 (1470) are used to form edges 1480, 1481 respectively in the etched strips of nitride layers.

4. Arsenic ion implantation with an ion dose of approximately  $5. \times 10^{15} \text{ cm}^{-2}$  is performed with an energy sufficient to penetrate oxide film 1475 and dope the surface in long strips of N+ doped regions (1402, 1401). Photoresist mask P.R.1 can be used as the mask for this step, but nitride layer 1474 can serve equally well as the implant mask. P.R.1 is stripped at the completion of this step.

5. An implant damage anneal and surface oxidation step follows, resulting in 200 to 300 nanometers of silicon dioxide 1462 grown over the source and drain diffusion strips. The temperature for this oxidation should be below 1000°C. to minimize the lateral diffusion of the N+ dopants in regions 1402, 1401. If desired it is possible through an extra masking step to remove nitride layer 1474 also from the field regions between adjacent channels, so as to grow oxide film 1462 not only over the source and drain regions but also over the field isolation regions.

6. In FIG. 14b a second photoresist mask P.R.2 (1482) is used to protect the source-side (1401) of the substrate during the subsequent implant step. This implant of boron can be performed at relatively high energy sufficient to penetrate through nitride layer 1474 and oxide layer 1475 but not high enough to penetrate top oxide 1473, nitride 1474 and oxide 1475. Alternatively, nitride layer 1474 can first be etched along edge 1482, using edge 1478 of the top oxide 1473 as a mask. The boron implant dose is in the range of  $1. \times 10^{13} \text{ cm}^{-2}$  and  $1. \times 10^{14} \text{ cm}^{-2}$ . The surface area of heavy p+ doping 1460 is confined to the very narrow and long strip of width extending between edge 1478 of the top oxide and the edge of the N+ diffusion 1402, and running the length of the drain diffusion strip. Note that the thick oxide 1462 prevents penetration of the boron implant into the drain diffusion strip. This greatly reduces the drain junction capacitance, which is highly desirable for fast reading. Note also that p+ region 1460 is automatically self aligned to drain region 1402 through this process.

7. Top oxide 1473, nitride 1474 and thin oxide 1475 are now removed by etching. This etching also reduces the thickness of the oxide layer 1462 protecting the source and drain diffusions. It is desirable to leave this film thickness at not less than approximately 100 nanometers at the completion of this etch step.

8. The remaining steps can be understood in relation to the structure of FIG. 14c: A gate oxide 1464 is grown over the surface, including the channel regions, separating between the long source/drain diffusion fusion strips (typical oxide thickness between 15 and 40 nanometers). A layer of polysilicon is deposited (thickness between 25 and 400 nanometers), doped N+, masked and etched to form continuous narrow strips of floating gates 1404 mask aligned to run parallel to drain diffusion strips 1402 and to overlap p+ regions 1460.

9. A second dielectric 1466, 1411 is grown or deposited on top of the substrate and floating gate strips, respectively. This can be a layer of silicon dioxide or a combination of thin films of silicon dioxide and silicon nitride, of combined thickness in the range between 20 and 50 nanometers.

10. A second layer of polysilicon is deposited, doped N+ (or silicided for lower resistivity), masked and etched to form control gates 1409 in long strips running perpendicular to the strips of floating gates and source/drain strips. Each control gate strip is capacitively coupled to the floating gate strips it crosses over through dielectric film 1411 in the areas where the strips overlap each other. Control gates 1409 also control the channel conduction in channel portions L2 not covered by the floating gate strips. Each strip of control gates is now covered by a dielectric isolation film (can be thermally grown oxide).

11. Using the strips of control gates as a mask, exposed areas of dielectric 1466, 1411 and of the strips of first polysilicon floating gates are etched away. The resulting structure has long strips, or rows, of control gates, each row overlying several floating gates 1404 where the outer edges of each floating gate are essentially self aligned to the edges defining the width of the control gate strip. These edges are now oxidized or covered with a deposited dielectric to completely insulate each floating gate. Field areas between adjacent rows of cells or between adjacent strips of source and drain regions are now automatically self aligned to the active device areas and do not require space consuming isoplanar oxidation isolation regions. (Of course, it is also possible to fabricate transistor 1400 with source, drain and channel regions defined by the edges of a thick isoplanar oxidation isolation layer, or to rely for field isolation on oxide 1462 grown also in the field regions, see the option described in step 5 above.)

The Eprom cell of this embodiment has several advantages over the prior art Eprom cells:

- a) Control gate 1409 now runs over a relatively thick oxide 1462 over the source and drain regions. Such a thick oxide is not possible for example with the prior art Eitan cell, where these source and drain regions are formed after, not before, the floating gate is formed. This improves the protection from oxide breakdowns and reduces the parasitic capacitance between control gate and drain.
- b) Control of parameter  $L_p$  and of the surface P+ doping concentration in region 1460 is superior to that afforded by the DSA prior art Tanaka cell.
- c) The device sensitivity to misalignment between floating gate and drain is far less than that experienced with the prior art Harari, Samachisa and Masuoka cells.
- d) For a given p+ concentration in the channel region, drain junction capacitance is less with this cell than with all other prior art devices, because p+ region 1460 is very narrowly confined near the drain diffusion.
- e) It is possible to dope p+ region 1460 to very high levels (which significantly enhances the programming efficiency) without unduly raising the conduction threshold voltage in the enhancement series channel region L2. This is particularly useful for Flash EEprom embodiments using this cell for the Eprom part. In such a Flash EEprom, the high initial threshold voltage in region  $L_p$  controlled by floating gate 1404 (initial  $V_t$  can be as high as +5.0V, the supply voltage, or higher), can be easily overcome by erasing the cell to lower threshold voltages. As an Eprom device the initial  $V_t$  in the unprogrammed state must not be higher than the control gate voltage during read, and this requirement sets an upper limit on how high the p+ doping concentration can be. Another limit on the magnitude of p+ doping concentration 1460 is established by the minimum drain voltage necessary for programming. The drain junction avalanche breakdown voltage must be at least as high as this minimum programming voltage.

## II. Self Aligned Split Channel Flash EEprom Cell With Isoplaner Field Isolation

FIG. 6a presents a topological view of a 2-times-2 memory array consisting of four Flash EEprom transistors 600a, 600b, 600c and 600d in accordance with one embodiment of this invention. FIG. 6b presents a cross section view of the same structure along AA of FIG. 6a. A second cross section along BB results in the Eprom transistor 500a shown in FIG. 5a.

Transistor 600a of FIG. 6a is a split channel Eprom transistor which has added to it erase gates 530, 535, which overlap edges 532a, 562a of floating gate 504a. Transistor 600a is programmed as a split channel Eprom transistor having a source diffusion 501a, a drain diffusion 502a, and a control gate 509. Floating gate 504a and channel portions L1 and L2 are formed in accordance

with the split channel Eprom transistor 500a of section I.a. or the split channel Eprom transistor 1400 of section I.b. However other split channel Eprom devices (such as the Eitan, Harari, Masuoka or Samachisa prior art Eprom) can also be used for the Eprom structure. The transistor channel width W is defined by the edges 505, 505a of a thick field oxide 562.

Transistor 600a is erased by tunneling of electrons from floating gate 504a to erase gates 530, 535, across tunnel dielectrics 531a, 561a on the sidewalls and top surface of the floating gate where it is overlapped by the erase gate.

Tunnel dielectric film 531a, 561a is normally a layer of Silicon Dioxide grown through thermal oxidation of the heavily N+ doped and textured polycrystalline silicon comprising the floating gate. It is well known in the industry (see for example an article by H.A.R. Wegener titled "Endurance Model for textured-poly floating gate memories", Technical Digest of the IEEE International Electron Device Meeting, December 1984, p. 480) that such a film, when grown under the appropriate oxidation conditions over properly textured doped polysilicon allows an increase by several orders of magnitude of the conduction by electron tunneling even when the film is several times thicker than tunnel dielectric films grown on single crystal silicon (such as the tunnel dielectric films used in the prior art Samachisa and Kynett devices). For example, a tunnel dielectric oxide grown to a thickness of 40 nanometers on N+ doped and textured polysilicon can conduct by electronic tunneling approximately the same current density as a tunnel dielectric oxide of 10 nanometers thickness grown on N+ doped single crystal silicon under identical voltage bias conditions. It is believed that this highly efficient tunneling mechanism is a result of sharp asperities at the grain boundaries of the polysilicon which is specially textured to enhance the areal density of such asperities. A commonly practiced technique is to first oxidize the surface of the polysilicon at a high temperature to accentuate the texturing, then stripping that oxide and regrowing a tunnel oxide at a lower temperature. The oxide film capping such an asperity experiences a local amplification by a factor of four to five of the applied electric field resulting in an efficient localized tunnel injector. The advantage provided by the thicker films of tunnel dielectric is that they are much easier to grow in uniform and defect-free layers. Furthermore the electric field stress during tunneling in the thick (40 nanometer) tunnel dielectric is only 25 percent of the stress in the thin (10 nanometer) tunnel dielectric, assuming the same voltage bias conditions. This reduced stress translates into higher reliability and greater endurance to write/erase cycling. For these reasons, all Flash EEprom embodiments of this invention rely on polypoly erase through a relatively thick tunnel dielectric.

In the embodiment of FIGS. 6a, 6b floating gate 504a is formed in a first layer of heavily N+ doped polysilicon of thickness between 25 and 400 nanometers, erase gates 530, 535 are formed in a second layer of N+ doped polysilicon of thickness between 50 and 300 nanometers, and control gate 509 is formed in a third conductive layer of thickness between 200 and 500 nanometers, which may be N+ doped polysilicon or a polycide, a silicide, or a refractory metal. The erase gate can be formed in a relatively thin layer because a relatively high sheet resistivity (e.g., 100 Ohm per square) can be tolerated since almost no current is carried in this gate during tunnel erase.

The manufacturing process can be somewhat simplified by implementing erase gates 530, 535 in the same conductive layer as that used for control gate 509. However the spacing Z between the edges of the control gate and the erase gate (and hence the cell size) would then have to be significantly greater than is the case when the control gate and erase gates are implemented in two different conductive layers insulated from each other by dielectric film 567a. In fact, in the triple layer structure 600a of FIG. 6a it is even possible to have control gate 509 slightly overlap one or both of the erase gates 530 and 535 (i.e., spacing Z can be zero or negative.) Transistor 600a employs a field isolation oxide 562 (FIG. 6b) of thickness between 200 and 1000 nanometers. Gate oxide 564a protecting channel portion L1 (512a) is thermally grown silicon dioxide of thickness between 15 and 40 nanometers. Dielectric film 567a which serves to

strongly capacitively couple control gate 509 and floating gate 504a is grown or deposited. It may be silicon dioxide or a combination of thin films of silicon dioxide and oxidized silicon nitride of combined thickness of between 20 and 50 nanometers. This dielectric also serves as part of the gate oxide protecting channel portion L2 (520a) as well as insulation 565a (FIG. 5a) over the source and drain diffusions. Erase dielectric 531a, 561a is thermally grown Silicon Dioxide or other deposited dielectrics possessing the appropriate characteristics for efficient erase conduction, such as Silicon Nitride. Its thickness is between 30 and 60 nanometers.

A point of significance is the fact that the tunnel dielectric area contributing to erase in each cell consisting of the combined areas of 531a and 561a, is insensitive to the mask misalignment between edges 532a, 562a of floating gate 504a and erase gates 530, 535. (Note that each erase gate, such as 530, is shared between two adjacent cells, such as 600a and 600c in this case). Any such misalignment will result in a reduction of the area of the tunnel dielectric at one edge of the floating gate, but also in an increase of equal magnitude in the area available for tunneling at the other edge of the floating gate. This feature permits the construction of a cell with very small area of tunnel dielectric. By contrast the prior art triple layer Flash EEprom cells of Masuoka and Kuo referenced above are sensitive to mask misalignment and therefore require a structure wherein the nominal area provided for tunnel erase may be much larger than the optimum such area, in order to accommodate for the worst case misalignment condition.

Another distinguishing feature of this embodiment relative to the Masuoka cell of FIGS. 3a and 3b is that Masuoka implements the erase gate in a first conductive layer 330 and the floating gate in a second conductive layer 304, i.e., in a reverse order to that used in this invention. This results in a far less efficient tunnel erase in the Masuoka cell because the asperities in Masuoka's tunnel dielectric 331 are at the surface of the erase gate (collector) rather than at the injecting surface of the floating gate. Therefore Masuoka's cell requires higher electric fields (and therefore higher V<sub>sub</sub>.ERASE voltages) than the structure of this invention.

Typical bias voltage conditions necessary to erase memory cells 600a, 600b, 600c and 600d are:

V<sub>sub</sub>.ERASE (on all erase gates 530, 535, 536)=15V to 25V applied for between 100 milliseconds and 10 seconds (the pulse duration is strongly dependent on the magnitude of V<sub>sub</sub>.ERASE), V<sub>sub</sub>.CG =0V, V<sub>sub</sub>.BB =0V. V<sub>sub</sub>.D and V<sub>sub</sub>.S can be held at 0V or at a higher voltage between 5V and 10V, so as to reduce the net voltage experienced during erase across dielectric film 565a in areas such as 563 (FIG. 6a) where erase gate 530 crosses over drain diffusion 502.

### III. Self Aligned Split Channel Flash EEprom Cell With Field Plate Isolation

A 2.times.2 array of Flash EEprom cells in accordance with another embodiment of this invention is shown in topological view in FIG. 7a and in two cross sectional views AA and CC in FIGS. 7b and 7c respectively. Cross sectional view BB is essentially the same as the split channel Eeprom transistor of FIG. 5a.

Split channel Flash EEprom transistor 700a employs three conductive layers (floating gate 704, erase gates 730, 735 and control gate 709) formed in the same sequence as described in section II in conjunction with the Flash EEprom transistor 600a of FIGS. 6a, 6b. The major distinguishing feature of transistor 700a is that erase gates 730, 735, 736 are used not only for tunnel erase but also as the switched off gates of isolation field transistors formed outside the active transistor regions. Thus, the thick isoplaner isolation oxide 562 of cell 600a (FIG. 6b) is not necessary, and is replaced inside the array of memory cells 700a, 700b, 700c and 700d by a much thinner oxide 762 (FIGS. 7b, 7c) capped with field plates 730, 735, 736, which are held at 0V at all times except during erasing.

The elimination of the thick isoplanar oxide inside the array of memory cells (this isoplanar oxide may still be retained for isolation between peripheral logic transistors) has several advantages:

1. The surface stress at the silicon-silicon dioxide boundary due to a prolonged thermal isoplanar oxidation cycle is eliminated inside the array, resulting in less leaky source and drain junctions and in higher quality gate oxides.
2. For a given cell width, the elimination of the isoplanar oxide allows the effective channel width  $W_{sub.1}$  under floating gate 704 to extend all the way between the two edges 732a, 762a of the floating gate. By comparison, effective channel width  $W$  of transistor 600a (FIG. 6b) is determined by the edges 505 of the isoplanar oxide and is therefore substantially smaller. This difference results in a higher read signal for cell 700a, or a narrower, smaller cell.
3. From capacitive coupling considerations (to be discussed in section VI below) the efficiency of tunnel erase is higher in cells where coupling of the floating gate to the silicon substrate 763 is greatest. In transistor 700a the entire bottom surface area of the floating gate is tightly coupled to the substrate 763 through the thin gate dielectric 764. By contrast, in transistor 600a (FIG. 6b) much of the bottom surface area of floating gate 504a overlies the thick field oxide 562 and is therefore not strongly capacitively coupled to substrate 563.
4. The width of control gate 709 between its edges 744 and 774 defines channel width  $W_{sub.2}$  of the series enhancement channel portion L2 (FIG. 7c). This permits the reduction in overall cell width due to removal of the requirement for the control gate to overlap the edges of the isoplanar oxide. One precaution necessary in the fabrication of cell 700a is that any misalignment between the mask layers defining edge 732a of floating gate 704a, edge 784 of erase gate 730, and edge 744 of control gate 709 must not be allowed to create a situation where a narrow parasitic edge transistor is created under control gate 709 in parallel with the split channel L1 and L2. However, as with cell 600a, since erase gates 730, 736 and control gate 709 are formed in two separate conductive layers which are isolated from each other by dielectric insulator film 767 (FIG. 7b) there is no requirement placed on the magnitude of the spatial separation  $Z$  between edge 784 and edge 744. In fact, the two edges can be allowed to overlap each other through oversizing or through misalignment, i.e.,  $Z$  can be zero or negative. Dielectric insulator 767 also forms part of the gate dielectric 766 (FIG. 7c) over channel portion L2.

In a memory array source diffusion 701 and drain diffusion 702 can be formed in long strips. If transistor 500a is used as the Eprom transistor, then source diffusion edge 721 is self aligned to the previously discussed sidewall spacer (not shown) while drain diffusion edge 723 is self aligned to edge 722 of floating gate 704a. In areas between adjacent floating gates 704a, 704c the source and drain diffusion edges (721x, 723x in FIG. 7a) respectively must be prevented from merging with one another. This can be accomplished by for example first forming floating gates 704a, 704c as part of a long continuous strip of polysilicon, then using this strip with an associated long continuous strip of sidewall spacer to form by ion implantation long diffusion strips 701, 702, removing the spacer strip, and only then etching the long continuous strip of polysilicon along edges 732a, 762a to form isolated floating gates 704a, 704c. As with the prior Flash EEprom embodiment it is possible to form this embodiment also in conjunction with Eprom cell 1400 (FIG. 14c) or with any other prior art split channel EEproms so long as they do not have their isoplanar isolation oxide inside the memory array.

#### IV. Self Aligned Split Channel Flash EEprom Cell with Erase Confined to The Vertical Edges of The Floating Gate

Another embodiment of the self aligned split channel Flash EEprom of this invention can result in a cell which has smaller area than cells 600a and 700a of the embodiments described in Sections

II and III respectively. In this third embodiment the area for tunnel erase between the floating gate and the erase gate is confined essentially to the surfaces of the vertical sidewalls along the two edges of each floating gate. To best understand how cell 800a of this embodiment differs from cell 700a a 2.times.2 array of cells 800a, 800b, 800c and 800d are shown in FIG. 8a in topological view and in FIG. 8b along the same cross section direction AA as is the case in FIG. 7b for cells 700a, 700c.

Cell 800a has a floating gate 804a formed in a first layer of heavily N+ doped polysilicon. This gate controls the transistor conduction in channel portion L1 (FIG. 8a) through gate oxide insulation film 864. Control gate 809 is formed in the second conductive layer, and is insulated from the floating gate by dielectric film 867, which may be a thermally grown oxide or a combination of thin silicon dioxide and silicon nitride films. Edges 874, 844 of control gate 809 are used as a mask to define by self aligned etching the edges 862a, 832a respectively of floating gate 804a. Erase gates 830, 835 are formed in a third conductive layer and are made to overlap edges 832a, 862a of floating gate 804a. Each erase gate such as 830 is shared by two adjacent cells (such as 800a, 800c).

The erase gates are insulated from control gate 809 by dielectric insulator 897 which is grown or deposited prior to deposition of erase gates 830, 835, 836. Tunnel erase dielectrics 831a, 861a are confined to the surface of the vertical edges 832a, 862a of the floating gate 804a. Erase gate 830 also provides a field plate isolation over oxide 862 in the field between adjacent devices.

The thickness of all conducting and insulating layers in structure 800 are approximately the same as those used in structure 700a. However, because the erase gate is implemented here after, rather than before the control gate, the fabrication process sequence is somewhat different. Specifically (see FIGS. 8a, 8b):

1. Floating gates 804a, 804c are formed in long continuous and narrow strips on top of gate oxide 864. The width of each such strip is L1 plus the extent of overlap of the floating gate over the drain diffusion.
2. Dielectric 867 is formed and the second conductive layer (N+ doped polysilicon or a silicide) is deposited.
3. Control gates 809 are defined in long narrow strips in a direction perpendicular to the direction of the strips of floating gates. The strips are etched along edges 844, 874, and insulated with relatively thick dielectric 897.
4. Edges 844, 874 (or the edges of insulator spacer 899 formed at both edges of control gate strip 809) are then used to etch dielectric 867 and then, in a self aligned manner to also etch vertical edges 832a and 862a of the underlying floating gate strips, resulting in isolated floating gates which have exposed edges of polysilicon only along these vertical walls.
5. Tunnel dielectric films 831a, 861a are formed by thermal oxidation of these exposed surfaces.
6. A third conductive layer is deposited, from which are formed erase gates 830 in long strips running in between and parallel to adjacent strips of control gates. These erase gates also serve as field isolation plates to electrically isolate between adjacent regions in the memory array.

Flash EEPROM transistor 800a can be implemented in conjunction with any of the split channel EEPROM transistors of this invention (transistors 500a and 1400) or with any of the prior art split gate EEPROM transistors of Eitan, Samachisa, Masuoka or Harari. For example, an array of Flash EEPROM transistors 800a can be fabricated by adding a few process steps to the fabrication

process for the split channel Eprom transistor 1400 (FIG. 14c), as follows:

Steps 1 through 10 are identical to steps 1 through 10 described in Section I.b. in conjunction with the manufacturing process for split channel Eprom transistor 1400.

Steps 11, 12, and 13 are the process steps 4, 5, and 6 respectively described in this section IV in conjunction with split channel Flash EEprom transistor 800a.

Cell 800a results in a very small area of tunnel erase, which is also relatively easy to control (it is not defined by a mask dimension, but rather by the thickness of the deposited layer constituting the floating gates). For this reason, this cell is the most highly scalable embodiment of this invention.

#### V. Self Aligned Split Channel Flash EEprom Cell With a Buried Erase Gate

A 2-times.2 array of Flash EEprom cells 900a, 900b, 900c and 900d in accordance with a fourth embodiment of this invention is shown in topological view in FIG. 9a and in two cross sectional views AA and DD in FIGS. 9b and 9c respectively. Cross section BB of FIG. 9a yields the split channel Eprom structure 500a of FIG. 5a.

Transistor 900a is a split channel Flash EEprom transistor having channel portions L1 and L2 formed by self alignment as in Eprom transistor 500a or in a non self aligned manner as in Eprom transistor 1400. Erase gate 930 is a narrow conductive strip sandwiched between floating gate 904a on the bottom and control gate 909 on top. Erase gate 930 is located away from edges 932a, 962a of the floating gate. These edges therefore play no role in the tunnel erase, which takes place through tunnel dielectric 931 confined to the area where erase gate 930 overlaps floating gate 904a. Erase gate 930 also overlaps a width W.sub.e of the series enhancement channel portion L2. During read or programming, erase gate 930 is held at OV, and therefore the channel portion of width W.sub.e does not contribute to the read or program current. The only contribution to conduction in channel portion L2 comes from widths W.sub.p and W.sub.q where the channel is controlled directly by control gate 909. Channel portion L1 however sees conduction contributions from all three widths, W.sub.p, W.sub.q and W.sub.e. Edges 932a, 962a of floating gate 904a can be etched to be self aligned to edges 944, 974 respectively of control gate 909. This then permits the formation of channel stop field isolation 998, by implanting a p type the field regions not protected by the control locating gate (FIG. 9b).

One advantage of cell 900a is that erase gate strips 930, 936 can be made very narrow by taking of controlled undercutting by for example etchings of the conductive layer forming these strips. This results in a small area of tunnel erase, which is insensitive to mask misalignment. Furthermore channel width W.sub.p and W.sub.q is also insensitive to mask misalignment. This embodiment of Flash EEprom can also be implemented in conjunction with prior art split channel Eproms cells such as the Eitan, Harari, Samachisa or Masuoka cells.

#### VI. Device Optimization

FIG. 10 represents a schematic of the major capacitances which couple the floating gate of the split channel Flash EEprom cells of this invention to the surrounding electrodes.

Specifically these are:

C.sub.G =Capacitance between Floating gate 1104 and control gate 1109.

C.sub.D =Capacitance between Floating gate 1104 and drain diffusion 1102.

C.sub.B =Capacitance between Floating gate 1104 and substrate 1163.

C.sub.E =Capacitance between Floating gate 1104 and erase gate 1130.

C.sub.T =C.sub.G +C.sub.D +C.sub.B +C.sub.E is the total capacitance. Q is the net charge stored on the floating gate. In a virgin device, Q=0. In a programmed device Q is negative (excess electrons) and in an erased device Q is positive (excess holes).

The voltage V.sub.FG on Floating gate 1104 is proportional to voltages V.sub.CG, V.sub.ERASE, V.sub.D, V.sub.BB and to the charge Q according to the following equation: ##EQU1##

In all prior art Eprom and Flash EEprom devices as well as in embodiment 600a of this invention, the dominant factor in C.sub.T is C. sub.G, the coupling to the control gate. However, in embodiments 700a, 800a and 900a C.sub.B is also a major contributor by virtue of the fact that the entire bottom surface of the floating gate is strongly coupled to the substrate.

#### a. Electrical Erase

During erase, the typical voltage conditions are V.sub.CG =OV, V. sub.D =OV, V.sub.S =OV, V.sub.BB =OV and V.sub.ERASE =20V. Therefore, substituting in equation(1),

$$V_{sub.FG} = Q/C_{sub.T} + 20C_{sub.E} /C_{sub.T} \quad (2)$$

The electric field for tunnel erase is given by

$$E_{sub.ERASE} = V_{sub.ERASE} /t - V_{sub.FG} /t \quad (3)$$

where t is the thickness of the tunnel dielectric. For a given V. sub.ERASE, E.sub.ERASE is maximized by making V.sub.FG small, which, from equation (2) is possible if C.sub.E /C.sub.T is small. Embodiments 700a, 800a and 900a allow this condition to be readily met: C.sub.E is small since the area of tunnel dielectric is small, and C.sub.T is large because both C.sub.G and C.sub.B are large. These embodiments are therefore particularly well suited for efficiently coupling the erase voltage across the tunnel dielectric.

#### b. Multistate storage

The split channel Flash EEprom device can be viewed as a composite transistor consisting of two transistors T1 and T2 in series— FIG. 11a. Transistor T1 is a floating gate transistor of effective channel length L1 and having a variable threshold voltage V.sub.T1. Transistor T2 has a fixed (enhancement) threshold voltage V.sub.T2 and an effective channel length L2. The Eprom programming characteristics of the composite transistor are shown in curve (a) of figure 11b. The programmed threshold voltage V.sub.tx is plotted as a function of the time t during which the programming conditions are applied. These programming conditions typically are V.sub.CG =12V, V.sub.D =9V, V.sub.S =V.sub.BB =OV. No programming can occur if either one of V.sub.CG or V. sub.D is at OV. A virgin (unprogrammed, unerased) device has V.sub.T1 =+ 1.5V and V.sub.T2 =+1.0V. After programming for approximately 100 microseconds the device reaches a threshold voltage V.sub.tx .gtreq.+6. 0 volts. This represents the off ("0") state because the composite device does not conduct at V.sub.CG =+5.0V. Prior art devices employ a so called "intelligent programming" algorithm whereby programming pulses are applied, each of typically 100 microseconds to 1 millisecond duration, followed by a sensing (read) operation. Pulses are applied until the device is sensed to be fully in the off state, and then one to three more programming pulses are applied to ensure solid programmability.

Prior art split channel Flash EEPROM devices erase with a single pulse of sufficient voltage  $V_{sub.ERASE}$  and sufficient duration to ensure that  $V_{sub.T1}$  is erased to a voltage below  $V_{sub.T2}$  (curve b) in FIG. 11b). Although the floating gate transistor may continue to erase into depletion mode operation (line (C) in FIG. 11b), the presence of the series T2 transistor obscures this depletion threshold voltage. Therefore the erased off ("1") state is represented by the threshold voltage  $V_{sub.tx} = V_{sub.T2} = +1.0V$ . The memory storage "window" is given by  $\Delta V = V_{sub.tx} ("0") - V_{sub.tx} ("1") = 6.0 - 1.0 = 5.0V$ . However, the true memory storage window should be represented by the full swing of  $V_{sub.tx}$  for transistor T1. For example, if T1 is erased into depletion threshold voltage  $V_{sub.T1} = -3.0V$ , then the true window should be given by  $\Delta V = 6.0 - (-3.0) = 9.0V$ . None of the prior art Flash EEPROM devices take advantage of the true memory window. In fact they ignore altogether the region of device operation (hatched region D in FIG. 11b) where  $V_{sub.T1}$  is more negative than  $V_{sub.T2}$ .

This invention proposes for the first time a scheme to take advantage of the full memory window. This is done by using the wider memory window to store more than two binary states and therefore more than a single bit per cell. For example, it is possible to store 4, rather than 2 states per cell, with these states having the following threshold voltage: ##EQU2## To sense any one of these four states, the control gate is raised to  $V_{sub.CG} = +5.0V$  and the source-drain current  $I_{sub.DS}$  is sensed through the composite device. Since  $V_{sub.T2} = +1.0V$  for all four threshold states transistor T<sub>sub.2</sub> behaves simply as a series resistor. The conduction current  $I_{sub.DS}$  of the composite transistor for all 4 states is shown as a function of  $V_{sub.CG}$  in FIG. 11c. A current sensing amplifier is capable of easily distinguishing between these four conduction states. The maximum number of states which is realistically feasible is influenced by the noise sensitivity of the sense amplifier as well as by any charge loss which can be expected over time at elevated temperatures. Eight distinct conduction states are necessary for 3 bit storage per cell, and 16 distinct conduction states are required for 4 bit storage per cell.

Multistate memory cells have previously been proposed in conjunction with ROM (Read Only Memory) devices and DRAM (Dynamic Random Access Memory). In ROM, each storage transistor can have one of several fixed conduction states by having different channel ion implant doses to establish more than two permanent threshold voltage states. Alternatively, more than two conduction states per ROM cell can be achieved by establishing with two photolithographic masks one of several values of transistor channel width or transistor channel length. For example, each transistor in a ROM array may be fabricated with one of two channel widths and with one of two channel lengths, resulting in four distinct combinations of channel width and length, and therefore in four distinct conductive states. Prior art multistate DRAM cells have also been proposed where each cell in the array is physically identical to all other cells. However, the charge stored at the capacitor of each cell may be quantized, resulting in several distinct read signal levels. An example of such prior art multistate DRAM storage is described in IEEE Journal of Solid-State Circuits, February 1988, p. 27 in an article by M. Horiguchi et al. entitled "An Experimental Large-Capacity Semiconductor File Memory Using 16-Levels/Cell Storage". A second example of prior art multistate DRAM is provided in IEEE Custom Integrated Circuits Conference, May 1988, p. 4.4.1 in an article entitled "An Experimental 2-Bit/Cell Storage DRAM for Macrocell or Memory-on-Logic Applications" by T. Furuyama et al.

To take full advantage of multistate storage in EPROMs it is necessary that the programming algorithm allow programming of the device into any one of several conduction states. First it is required that the device be erased to a voltage  $V_{sub.T1}$  more negative than the "3" state (-3.0V in this example). Then the device is programmed in a short programming pulse, typically one to ten microseconds in duration. Programming conditions are selected such that no single pulse can shift the device threshold by more than one half of the threshold voltage difference between two successive states. The device is then sensed by comparing its conduction current  $I_{sub.DS}$  with that of a reference current source  $I_{sub.REF}$ ,  $i$  ( $i=0,1,2,3$ ) corresponding to the desired conduction state (four distinct reference levels must be provided corresponding to the four

states). Programming pulses are continued until the sensed current (solid lines in FIG. 11c) drops slightly below the reference current corresponding to the desired one of four states (dashed lines in FIG. 11c). To better illustrate this point, assume that each programming pulse raises  $V_{sub.tx}$  linearly by 200 millivolts, and assume further that the device was first erased to  $V_{sub.T1} = 3.2V$ . Then the number of programming/sensing pulses required is: ##EQU3## In actual fact shifts in  $V_{sub.tx}$  are not linear in time, as shown in FIG. 11b (curve (a)), therefore requiring more pulses than indicated for states "1" and "0". If 2 microseconds is the programming pulse width and 0.1 microseconds is the time required for sensing, then the maximum time required to program the device into any of the 4 states is approximately  $39 \times 2 + 39 \times 0.1 = 81.9$  microseconds. This is less than the time required by "intelligent programming algorithms" of prior art devices. In fact, with the new programming algorithm only carefully metered packets of electrons are injected during programming. A further benefit of this approach is that the sensing during reading is the same sensing as that during programming/sensing, and the same reference current sources are used in both programming and reading operations. That means that each and every memory cell in the array is read relative to the same reference level as used during program/sense. This provides excellent tracking even in very large memory arrays.

Large memory systems typically incorporate error detection and correction schemes which can tolerate a small number of hard failures i. e. bad Flash EEPROM cells. For this reason the programming/sensing cycling algorithm can be automatically halted after a certain maximum number of programming cycles has been applied even if the cell being programmed has not reached the desired threshold voltage state, indicating a faulty memory cell.

There are several ways to implement the multistate storage concept in conjunction with an array of Flash EEPROM transistors. An example of one such circuit is shown in FIG. 11e. In this circuit an array of memory cells has decoded word lines and decoded bit lines connected to the control gates and drains respectively of rows and columns of cells. Each bit line is normally precharged to a voltage of between 1.0 V and 2.0 V during the time between read, program or erase. For a four state storage, four sense amplifiers, each with its own distinct current reference levels  $I_{REF,0}$ ,  $I_{REF,1}$ ,  $I_{REF,2}$ , and  $I_{REF,3}$  are attached to each decoded output of the bit line. During read, the current through the Flash EEPROM transistor is compared simultaneously (i.e., in parallel) with these four reference levels (this operation can also be performed in four consecutive read cycles using a single sense amplifier with a different reference applied at each cycle, if the attendant additional time required for reading is not a concern). The data output is provided from the four sense amplifiers through four Di buffers ( $D_0$ ,  $D_1$ ,  $D_2$  and  $D_3$ ).

During programming, the four data inputs  $I_i$  ( $I_0$ ,  $I_1$ ,  $I_2$  and  $I_3$ ) are presented to a comparator circuit which also has presented to it the four sense amp outputs for the accessed cell. If  $D_i$  match  $I_i$ , then the cell is in the correct state and no programming is required. If however all four  $D_i$  do not match all four  $I_i$ , then the comparator output activates a programming control circuit. This circuit in turn controls the bit line (VPBL) and word line (VPWL) programming pulse generators. A single short programming pulse is applied to both the selected word line and the selected bit line. This is followed by a second read cycle to determine if a match between  $D_i$  and  $I_i$  has been established. This sequence is repeated through multiple programming/reading pulses and is stopped only when a match is established (or earlier if no match has been established but after a preset maximum number of pulses has been reached).

The result of such multistate programming algorithm is that each cell is programmed into any one of the four conduction states in direct correlation with the reference conduction states  $I_{sub.REF}$ ,  $i$ . In fact, the same sense amplifiers used during programming/reading pulsing are also used during sensing (i.e., during normal reading). This allows excellent tracking between the reference levels (dashed lines in FIG. 11c) and the programmed conduction levels (solid lines in FIG. 11c), across large memory arrays and also for a very wide range of operating temperatures.

Furthermore, because only a carefully metered number of electrons is introduced onto the floating gate during programming or removed during erasing, the device experiences the minimum amount of endurance-related stress possible.

In actual fact, although four reference levels and four sense amplifiers are used to program the cell into one of four distinct conduction states, only three sense amplifiers and three reference levels are required to sense the correct one of four stored states. For example, in FIG. 11c,  $I_{sub.REF}$  ("2") can differentiate correctly between conduction states "3" and "2",  $I_{sub.REF}$  ("1") can differentiate correctly between conduction states "2" and "1", and  $I_{sub.REF}$  ("0") can differentiate correctly between conduction states "1" and "0". In a practical implementation of the circuit of FIG. 11e the reference levels  $I_{sub.REF}$ ,  $i$  ( $i=0,1,2$ ) may be somewhat shifted by a fixed amount during sensing to place them closer to the midpoint between the corresponding lower and higher conduction states of the cell being sensed.

Note that the same principle employed in the circuit of FIG. 11e can be used also with binary storage, or with storage of more than four states per cell. Of course, circuits other than the one shown in FIG. 11e are also possible. For example, voltage level sensing rather than conduction level sensing can be employed.

#### c. Improved Charge Retention

In the example above, states "3" and "2" are the result of net positive charge (holes) on the floating gate while states "1" and "0" are the result of net negative charge (electrons) on the floating gate. To properly sense the correct conduction state during the lifetime of the device (which may be specified as 10 years at 125.degree. C.) it is necessary for this charge not to leak off the floating gate by more than the equivalent of approximately 200 millivolts shift in  $V_{sub.T1}$ . This condition is readily met for stored electrons in this as well as all prior art Eprom and Flash EEPROM devices. There is no data in the literature on charge retention for stored holes, because, as has been pointed out above, none of the prior art devices concern themselves with the value  $V_{sub.T1}$  when it is more negative than  $V_{sub.T2}$ , i.e., when holes are stored on the floating gate. From device physics considerations alone it is expected that retention of holes trapped on the floating gate should be significantly superior to the retention of trapped electrons. This is because trapped holes can only be neutralized by the injection of electrons onto the floating gate. So long as the conditions for such injection do not exist it is almost impossible for the holes to overcome the potential barrier of approximately 5.0 electronvolts at the silicon-silicon dioxide interface (compared to a 3.1 electron volts potential barrier for trapped electrons).

Therefore it is possible to improve the retention of this device by assigning more of the conduction states to states which involve trapped holes. For example, in the example above state "1" had  $V_{sub.T1} = +2.0V$ , which involved trapped electrons since  $V_{sub.T1}$  for the virgin device was made to be  $V_{sub.T1} = +1.5V$ . If however  $V_{sub.T1}$  of the virgin device is raised to a higher threshold voltage, say to  $V_{sub.T1} = +3.0V$  (e.g. by increasing the p-type doping concentration in the channel region 560a in FIG. 5a), then the same state "1" with  $V_{sub.T1} = +2.0V$  will involve trapped holes, and will therefore better retain this value of  $V_{sub.T1}$ . Of course it is also possible to set the reference levels so that most or all states will have values of  $V_{sub.T1}$  which are lower than the  $V_{sub.T1}$  of the virgin device.

#### d. Intelligent Erase for Improved Endurance

The endurance of Flash EEPROM devices is their ability to withstand a given number of program/erase cycles. The physical phenomenon limiting the endurance of prior art Flash EEPROM devices is trapping of electrons in the active dielectric films of the device (see the Wegener article referenced above). During programming the dielectric used during hot electron

channel injection traps part of the injected electrons. During erasing the tunnel erase dielectric likewise traps some of the tunneled electrons. For example, in prior art transistor 200 (FIG. 2) dielectric 212 traps electrons in region 207 during programming and in region 208 during erasing. The trapped electrons oppose the applied electric field in subsequent write/erase cycles thereby causing a reduction in the threshold voltage shift of  $V_{sub.tx}$ . This can be seen in a gradual closure (FIG. 11d) in the voltage "window" between the "0" and "1" states of prior art devices. Beyond approximately  $1 \times 10^{sup. 4}$  program/erase cycles the window closure can become sufficiently severe to cause the sensing circuitry to malfunction. If cycling is continued the device eventually experiences catastrophic failure due to a ruptured dielectric. This typically occurs at between  $1 \times 10^{sup. 6}$  and  $1 \times 10^{sup. 7}$  cycles, and is known as the intrinsic breakdown of the device. In memory arrays of prior art devices the window closure is what limits the practical endurance to approximately  $1 \times 10^{sup. 1}$ .

(19)日本国特許庁 (JP)

## (12)特許公報 (B2)

(11)特許番号

第2694618号

(45)発行日 平成9年(1997)12月24日

(24)登録日 平成9年(1997)9月12日

(51)Int.Cl. <sup>6</sup>	識別記号	府内整理番号	F I	技術表示箇所
H 01 L 21/8247			H 01 L 29/78	3 7 1
27/115			27/10	4 3 4
29/788				
29/792				

請求項の数22(全 27 頁)

(21)出願番号	特願平1-146050	(73)特許権者	999999999 エリヤホウ ハライ アメリカ合衆国 95030 カリフォルニア アロス ガトス, アウゼライス コート 104
(22)出願日	平成1年(1989)6月8日	(72)発明者	エリヤホウ ハライ アメリカ合衆国 95030 カリフォルニア アロス ガトス, アウゼライス コート 104
(65)公開番号	特開平2-110981	(74)代理人	弁理士 井ノ口 寿
(43)公開日	平成2年(1990)4月24日	審査官	河合 章
(31)優先権主張番号	204175	(56)参考文献	特開 平2-1987 (JP, A) 英国公開2200795 (GB, A) 欧州公開193841 (EP, A1) 欧州公開164781 (EP, A2)
(32)優先日	1988年6月8日		
(33)優先権主張国	米国 (US)		

(54)【発明の名称】 フラッシュEEPROMセルアレイとそのアレイの形成方法

## (57)【特許請求の範囲】

【請求項1】半導体基板の表面上にフラッシュEEPROMセルのアレイを形成する方法において：

前記基板表面を横切って、前記基板表面の下に第1の方向に伸びて第1の方向とほぼ直角である第2の方向に間隔を保って設けられその間にチャンネル領域を形成するための平行な複数のソースおよびドレイン領域を形成するステップと、

前記基板から絶縁され、前記第1の方向に伸びる連続する長さをもちかつ前記第2の方向に間隔を保って形成されるチャンネル領域のすくなくとも一部を越えて配置され、平行に伸びている複数の第1の導電性の帯条を形成するステップと、

前記基板および前記第1の導電性の帯条と絶縁され、前記第2の方向に連続して伸びかつ前記第1の方向に対し

て間隔を保って形成されたコントロールゲートとしての複数の第2の導電性の帯条を形成するステップと、前記第2の帯条の縁を基準として前記第2の帯条の間に空間を保って形成されたマスク開口を介して、前記第1の帯条の一部を除去し、前記第2の帯条の間にフローティングゲートの領域の表面を露出させるステップと、前記露出させられたフローティングゲートの表面領域にトンネル誘電体層を形成するステップと、および前記第2の帯条と前記基板から絶縁された状態で前記第1の方向に間隔を保ち、前記第1の方向に連続的に伸びて消去ゲートを形成するための第3の複数の平行に伸びる導電性帯条を前記トンネル誘電体層のそれぞれに接触して前記第2の帯条間に形成するステップとからなるフラッシュEEPROMの製造方法。

【請求項2】請求項1記載の方法において、前記第1の

帯条の一部を除去するステップは隣接する第2の帯条間の空間を実質的に完全に伸びるマスクを開口を介して行うエッティングを含むフラッシュEEPROMの製造方法。

【請求項3】請求項1記載の方法において、前記第1の帯条の一部を除去するステップは前記第1の帯条を前記第2の隣接する帯条間の開口によってエッティングされるのであるが、その開口は前記間隔中の前記第2の帯条に沿って誘電体のスペーサを形成することによって減少させられるフラッシュEEPROMの製造方法。

【請求項4】請求項1記載の方法において、前記第1の帯条を形成するステップは前記第2の方向のチャンネル領域に限られる幅の帯条の形成を含むフラッシュEEPROMの製造方法。

【請求項5】請求項1記載の方法において、前記第3の帯条を形成するステップは前記コントロールゲート間のそれぞれの間隔の中に前記第3の帯条を形成するステップを含むフラッシュEEPROMの製造方法。

【請求項6】請求項1記載の方法において、前記第3の帯条の形成は前記第3の帯条の各々を前記第2の帯条間の空間の両サイド上のフローティングゲートの表面領域に支持されている誘電体層に接触させる工程を含むフラッシュEEPROMの製造方法。

【請求項7】請求項1記載の方法において、前記第3の帯条を形成するステップは、前記各々の第3の帯条の幅を前記第2の方向に前記第2の帯条の両側に存在するように形成する工程を含むフラッシュEEPROMの製造方法。

【請求項8】請求項1記載の方法において、前記第1の帯条を前記第2の帯条の間で除去するステップはそれらの間の第1の帯条の全部よりも少ない部分を除去することによって前記第2の帯条間にフローティングゲート領域を残すフラッシュEEPROMの製造方法。

【請求項9】請求項8記載の方法において、前記フローティングゲートの露出された表面上に誘電体層を形成するステップは前記誘電体層を前記第2の帯条間の間隔中に伸びるフローティングゲートの表面上に誘電体層を形成する工程を含むフラッシュEEPROMの製造方法。

【請求項10】請求項1記載の方法において、前記ソースとドレイン領域を基板中に形成する工程は前記第1の帯条を形成した後で、前記第1の帯条間に形成されたマスク開口を用いて前記第1の帯条の縁を参照して行われるフラッシュEEPROMの製造方法。

【請求項11】請求項10記載の方法において、前記マスク開口は前記第1の帯条の縁に沿って形成された誘電体スペーサを用いて形成されるフラッシュEEPROMの製造方法。

【請求項12】請求項1ないし11のいずれか1つにしたがう方法であって、前記第1、第2および第3の帯条は前記アレイの長さに沿って実質的に均一な幅に沿って直線的に形成されるものであるフラッシュEEPROMの製造方法。

【請求項13】請求項1ないし11記載のいずれかの方法にしたがう方法であって、前記第1、第2および第3の導電体帯条を形成するステップは、導電性をもつようにドープされたポリシリコン材料によって形成されるそれぞれ異なる第1、第2および第3の層を形成するものであるフラッシュEEPROMの製造方法。

【請求項14】半導体基板表面上に形成される高密度フラッシュEEPROMセルアレイにおいて：

連続的に伸びた導電性にドープされ複数の領域であり、前記基板内に形成され、第1の方向に実質的に直線経路に伸びた長さをもっており、前記ドープされた領域はソースとドレイン領域を形成し、そして前記基板表面を横切って第2の方向に分離され、その間にチャンネル領域を形成し、前記第1および第2の方向は実質的にそれぞれ直角である連続的に伸びた導電性にドープされた複数の領域と、

矩形のフローティングゲートの2次元アレイであり、実質的に直線上の列に配置され、チャンネル領域に沿って第1の方向に伸びており、かつ実質的に直線上に第2の方向に伸びている行に配列されている矩形のフローティングゲートの2次元アレイと、

複数の長いコントロールゲートは、前記第2の方向に直線的に伸び、そして前記第1の方向の間隔を保っており、各々の前記コントロールゲートは前記フローティングゲートの行の1つに対応してその上にあり、その間の電気的絶縁を保ち、そして、前記フローティングゲート表面が前記コントロールゲート間の空間に対面しており複数の長いコントロールゲートと、

前記フローティングゲートの表面に支持されるトンネル誘電体層と、および

複数の長い消去ゲートであり、前記第2の方向にのび実質的に直線経路にのびており、そして前記コントロールゲート間の空間に第1の方向に位置させられ、前記フローティングゲートの表面により支持されるトンネル誘電体層に接触するように位置させられている複数の長い消去ゲートと、

からなるフラッシュEEPROM装置。

【請求項15】請求項14記載のフラッシュEEPROMセルアレイにおいて、前記フローティングゲートは前記チャンネル領域の部分の一部の上に位置させられ、前記第2の方向にドープされた近接する領域の間に位置させられることにより、スプリットチャンネルセルのアレイを形成するフラッシュEEPROM装置。

【請求項16】請求項14記載のフラッシュEEPROMセルアレイにおいて、前記消去ゲートは前記第1の方向のコントロールゲート間の間隔の各々に実質的に位置させられているフラッシュEEPROM装置。

【請求項17】請求項16記載のフラッシュEEPROMセルアレイにおいて、

前記各々のフローティングゲートは前記コントロールゲ

一ト間の空間に向かう前記フローティングゲートの反対側の2つの表面を持ち、そこにおいて、前記トンネル誘電体は前記フローティングゲートの反対側の表面で支持されており、異なる消去ゲートと接触しているフラッシュ EEPROM装置。

【請求項18】請求項14記載のフラッシュ EEPROMセルアレイにおいて、前記フローティングゲートの表面は前記隣接するコントロールゲート間の空間の内側に延びているフラッシュ EEPROM装置。

【請求項19】請求項18記載のフラッシュ EEPROMセルアレイにおいて、前記フローティングゲートの前記表面は、前記空間内の隣接する前記フローティングゲートの頂面の上のある領域を含むフラッシュ EEPROM装置。

【請求項20】請求項14記載のフラッシュ EEPROMセルアレイにおいて、前記消去ゲートは前記第1の方向で、隣接するコントロールゲートの方に余分にその間に絶縁を保って延びる幅もつ、フラッシュ EEPROMセルアレイ。

【請求項21】請求項14記載のフラッシュ EEPROMセルアレイにおいて、前記消去ゲートは前記制御ゲート間の個々のスペースの両側のトンネルダイオードの誘電体層に接触しているフラッシュ EEPROM装置。

【請求項22】請求項14ないし21のいずれか1つのEEPROMセルアレイにおいて、前記フローティング、コントロールおよび消去ゲートは導電性にドープされたポリシリコン材料の別々の層から形成されているフラッシュ EEPROM装置。

#### 【発明の詳細な説明】

##### (産業上の利用分野)

この発明は一般的には電気的にプログラム可能なリードオンリーメモリ (Eeprom) 半導体と、電気的に消去可能でプログラム可能なリードオンリーメモリ (EEeprom) 、さらに詳しく言えばそのようなメモリの構造とそれを製造する工程に関する。

##### (発明の背景)

電気的にプログラム可能なリードオンリーメモリ (Eeprom) は、フィールドフェクトトランジスタ構造で、半導体基板領域のチャンネルから絶縁されてソースとドレイン領域間に設けられているフローティング導通ゲート（無接続）を用いている。コントロールゲートはフローティングゲートの上に設けられており、それから絶縁されている。

そのトランジスタのしきい値電圧特性はそのフローティングゲート上に引き留められる電荷の量によってコントロールされる。すなわち、そのソースとドレイン領域間の導通を許容するために、トランジスタがターンオンされる前にそのコントロールゲートに加えられなくてはならない電圧、すなわちその電圧がしきい値電圧である。

トランジスタは、その基板のチャンネル領域の薄い誘

電体のゲートを通してフローティングゲートに電子を加速することによって、2つの状態のうちの1つをプログラムすることができる。

メモリセルのトランジスタの状態は、そのトランジスタのソースとドレインとコントロールゲートに動作電圧を与えることによって読むことができ、それから制御ゲート電圧が選択された時にソースとドレイン間を流れる電流を検出することによりその装置がオンにプログラムされているか、またはオフにプログラムされているかを知ることができる。

Eepromセルの二次元アレイの中の特定の1つのセルを読み出しのためにはアドレスするためには、そのセルがアドレスされるべきセルが含まれているコラムのソースとドレイン線間にソースとドレイン電圧を与えること、およびアドレスされるべきセルが含まれている行のコントロールゲートにコントロール電圧を与えることによりなされる。

そのようなメモリセルの例として三重ポリシリコン、チャンネル分割形電気的消去可能かつプログラム可能なリードオンリーメモリ (Eeprom) がある。フローティングゲートとコントロールゲートがチャンネルの近接部分上に延びているので、これはスピリットチャンネル装置と言われている。これにより、トランジスタ構造は直列の2つのトランジスタとして働き、その1つはフローティングゲート上の電荷レベルに応答する変化しきい値を持ち、他の1つはそのフローティングゲートの電荷には影響されないで、むしろ通常のフィールドフェクトトランジスタと同様にそのコントロールゲートに印加される電圧に応答して働く。

そのようなメモリセルは三重ポリシリコンセルといわれている。なぜならそれは、ポリシリコン材料の三重の導電層をもっているからである。フローティングとコントロールゲートに加えるにさらに消去ゲートが含まれている。消去ゲートは各メモリセルトランジスタのフローティングゲート表面に近接して通過しているが、それから薄いトンネル誘電体（トンネル効果を持つ）によって絶縁されている。

セルのフローティングゲートから電荷が消去ゲートにすべてのトランジスタに適当な電圧が印加されたときに除去される。セルの全体のアレイまたは特別のセルのグループが同時に消去（すなわちフラッシュによって）されるときに、そのようなEepromのセルをフラッシュEepromアレイという。

本発明の第1の目的は、セルを小形にして、かつ、チップ状の集積密度を増大させることができるEepromとEEepromセルのアレイ構造とその製造方法を提供することにある。

本発明のさらに他の目的は、その構造が製造に適しており、信頼性があり、計測可能であり、再現性があり、高い収率で製造可能である構造を提供することにある。

本発明のさらに他の目的は、磁気ディスク記憶装置に置き換えることができる固体メモリとして使用することができるEEpromの半導体チップを提供することにある。

(発明の要約)

これら、およびさらに付加的な目的は、本発明の種々の局面から、またはそれらを組み合わせることにより達成されるものであり、主たる本発明の側面を以下に記述する。

本発明の1つの側面によれば、EepromまたはフラッシュEEprom装置の分割チャンネルのフローティングゲートの一方の端はドレイン拡散領域の端に自己整合、または自己整合されそして重ね合わされており、第2のフローティングゲートの端はソース拡散領域の端に自己整合されているが、そこから離れている。

ソースに対面する第2のフローティングゲートの第2の端に沿って形成される側壁のスペーサは2つの端面間の間隔の程度を定義するのに用いられている。

フローティングゲートの端に対するソースとドレインの自己整合は最も重要な3つの装置定数を正確に制御することができるチャンネル分割形のEeprom装置が得られるものであり、チャンネルセグメントの長さL1とL2はフローティングゲートとコントロールゲートによってそれぞれ制御され、また、フローティングゲートとドレインの拡散ゲート間の重なり合いの程度によってもしかりである。

すべての3つのパラメータはマスクの不一致に対してあまり敏感ではなく、小さくスケールダウンされて作られた装置においても、再現可能である。

本発明の他の局面によれば、ドレインディフージョン領域に近接するチャンネルの濃くドープされた領域は新規なしかもよく制御できる技術によって形成される。

このチャンネル領域の長さLpとドーピングの密度はプログラミングと読み出しのための主要なパラメータになるものであり、これにより、フローティングゲートとソースまたはドレイン領域の不一致に対して比較的敏感でないチャンネル構造の形成を許容するものである。

1つのEeprom装置をフラッシュEeprom装置に転換するために分離された消去ゲートが設けられる。

本発明のさらに他の側面によれば、フローティングゲートと消去ゲート間の重ね合わせ領域はマスクの位置合わせずれに対して敏感でなく、それ故に大変に小さく作ることができる。

本発明の特定の形態によれば、消去ゲートは近接するメモリセルの中で隣接するセル間の分離の電気的な非常に小さな分離帯として働くフィールドプレートとして利用されることができる。

本発明のさらに他の側面によれば、大変に薄いフローティングゲートのエッジを仕上げることを用いることによって、その電子注入手段としての効果を増大させるようする消去メカニズムが提供される。

これらの種々の特徴を結合することによって、新しいチャンネル分割のEepromやチャンネル分割形のフラッシュEEprom装置を作り出すことができ、それらは非常に製造が容易であり、スケールダウンすることができるし、より大きな蓄積密度をもち、さらに信頼性も高いのである。すなわち、それらは従来存在しているEepromとかフラッシュEEpromに比べてそうなのである。

本発明のいろいろな局面を利用したメモリは特にコンピュータシステムにおいて、現存している磁気貯蔵媒体（ハードディスクとかフロッピーディスク）を置き換えるのに非常に有用である。

なぜならば、そこに貯蔵されるであろう情報の密度が大変に高いからである。

さらに、他の付加的な目的とか、この発明の利点は好適な実施例の記述により理解されるべきであり、その好適な実施例は添付された図面を参照して説明される。

(先行技術の説明)

三重ポリシリコン装置は、ジェイ. クペック (J. Kupre c) 等により1980年のアイイーディーエムテクニカルダイジェスト (IEDM Technical Digest) の602頁に「1ビットあたり1つのトランジスタを用いた3レベルのポリシリコンEEprom」と題した論文に記述されていた。

クペックの装置に対する改良は、エフ. マスオカ (F. Masuoka) とエイチ. イイヅカ (H. Iizuka) が1985年7月23日に発行された米国特許第4,531,203号に提案された。

同じセルに対しての変形は、1985年の12月24日に発行された米国特許第4,561,004号にシーケー. クオ (C. K. Kuo) とエス. シー. ツアール (S. C. Tsaur) によって記述されており、また、エフ. マスオカ (F. Masuoka) 等によりアイイーディー (IEEE) の国際半導体回路会議、1985年2月の168頁「三重ポリシリコン技術を用いた256KのフラッシュEEprom」と題する論文にも示されている。

クペックの装置の断面図は第1図に示されている。

トランジスタ200bは積み重ねられたフローティングゲート204bと制御ゲート209bをもっており、制御ゲート209bはソース201bとドレイン202bがフローティングゲートの端に自己整合されて設けられている。

ゲートの誘電体212は比較的厚いので、フローティングゲートからソースまたはドレインへのトンネル消去を許容しない。消去板230bがコントロールゲートの上にあり、コントロールゲートフローティングゲートの両側壁を覆っている。消去はフローティングゲート204bと消去プレート230b間の比較的薄い酸化物層231bを通過するトンネル現象により行われる。

クペックは過消去の条件を克服する企として消去プレートを高い電圧の消去の間にドレイン202bと第1図に示されていない高いインピーダンスの抵抗Rを介して消去電圧V<sub>ERASE</sub>に接続することにより行った。

そのセルが空乏層状態に消去されるとドレインからソースへのトランジスタの導通電流は低下し、消去電圧の殆どは抵抗端子間に現れ、消去板230bの電圧をトンネル電圧以下に減少させる。

このやり方は大きなアレイで部分的な消去を行う場合には非常に困難である。

なぜならば、異なった時に異なってトランジスタが導通を開始するからである。

第2図はマスオカ等の先行技術の素子の平面図を示している。

トランジスタ300はチャンネル分離形のEepromトランジスタから成り立っており、そのEepromトランジスタはソース301、ドレイン302、チャンネルの断面L1(312)に沿うチャンネル導通を制御するフローティングゲート304、前記フローティングゲートに容量的に結合しており、また、エンハンスマントのしきい値電圧をもつ値L(320)の直列部分に沿う導通を制御するコントロールゲート300をもっている。

このトランジスタのチャンネル幅(W)は、これはソースとドレイン拡散領域のエッジと同じようにイソブレーナ酸化により形成される薄い酸化膜の端305によって規定される。

25から40ナノメータの厚さをもっている酸化膜層332は、フローティングゲートと基板の分離のために用いられている。

マスオカはフローティングゲートの端に沿う消去ゲート330を追加した。この消去ゲートはフローティングゲートと消去ゲートが重なっているトンネル誘電体の領域におけるフローティングゲート304を電気的に消去するために用いられる。トンネル誘電体331は、30から60ナノメータの間の厚さをもっている。

マスオカは消去期間に以下の電圧を用いることを規定している。

$V_s = 0V$ ,  $V_d = 0V$ ,  $V_{ce} = 0V$ ,  $V_{bb} = 0V$ ,  $V_{erase} = 20V$ から30V

マスオカの素子の重なり合う領域331は、この重なり合いを規定する2つのマスク間の合わせ誤差に対して鋭敏である。

重なり合う領域は非常に小さいから、僅かな合わせ誤差であっても、トンネル消去のために用いられるこの領域の変化に大きな変化を与えることになる。

この変化は、ウエーハごとに大変な誤差となって現れる。

Eepromのプログラムの挙動を理想化するために2つの明確に別れたチャンネル領域をもつEepromの他の先行技術は、エス・タナカ(S. Tanaka)等により1984年のアイエスエスシーシー(ISSCC)の技術ペイパーのダイジェストの第148頁に「チップ上にテスト回路をもつ256Kのプログラム可能なシーモスのEeprom」(A Programmable 256K CMOS Eeprom with On Chip Test Circuits)と題する記事に示されている。

この装置の断面図が第3図に示されている。これはタナカの論文の第3図に対応するものである。トランジスタ400eは積み重ねゲートのEepromトランジスタであり

(分離チャンネル形ではない)、フローティングゲート404eとコントロールゲート409eの両エッジに自己整合されたソース401eとドレイン402eをもっている。チャンネル領域はP基板463eよりも多くドープされているが、しかしその領域460eよりも多くP形にドープされた第2のp+領域477eが設けられている。

この447eの領域はドレイン側の表面からのみのポロンの下または横方向への拡散により形成され、ドレイン側のフローティングゲートに自己整合するようフローティングゲートの製造以後に形成される。

N+のドレイン接合領域からの砒素の横方向拡散に先立つポロンの横方向の拡散の程度がドレインに隣接するチャンネル領域Lp(478e)を規定する。

これはDMOSタイプ構造であり、タナカによつてはDSA(拡散自己整合)と呼ばれている。

p+領域478eの存在は、高電圧によるプログラム過程におけるドレイン空乏層領域の幅をかなり減少させる。

短い空乏層の幅は空乏層領域に入るチャンネル電子に対してより多くのエネルギーを加えることになる。そのことは、ホットエレクトロンの注入によるプログラム効率にかなりの向上させる結果として現れる。

トランジスタ400eの製造は、困難であることは証明された。なぜならば、それは二重の拡散工程によって長さLpと表面チャンネルのp+濃度を制御することが困難であるからである。

加えるに拡散より、0.3ミクロンより大きなLpの値を得ることが困難である。なぜならば、この構造の規格のスケールでは、かなり低い温度拡散サイクルの資料が要求されるからである。

さらに、このDSAEeprom装置はプログラムされていない領域(導通領域)において過度な高いしきい値電圧を要求すると同時に、高いドレイン接合容量を要求するという困難がある。

これらの2つの影響により実質的に読みのアクセス時間が増大させられる。

(実施例の詳細な説明)

I. a. 自己整合形のドレイン拡散と自己整合形の離れたソース拡散をもつ分離チャンネルのEepromトランジスタ

第4a図は、本発明の第1の実施例に従うところの分離チャンネル形のEepromトランジスタの断面図である。

トランジスタ500aはp形のシリコン基板563(これはp++にドープされたシリコンサブストレイトの表面にエピタキシャル成長法でpタイプのものに置き換えることができる), N+のソース拡散領域501a, N+のドレイン拡散領域502a, 周りの基板よりも多くドープされたチャンネル領域560a, チャンネル512aのL1の部分に存在するフローティングゲート504a, チャンネル520aのL2の残さ

れた部分でありフローティングゲートの部分であるところに存在するコントロールゲート509からできている。フローティングゲート504aはシリコン基板の表面から誘電体フィルム564aによって誘電体的に分離されている。その誘電体は熱的に生長させられたシリコン酸化物である。

コントロールゲート509はフローティングゲート504aに誘電体フィルム567を介して容量的に結合させられている。そして、この誘電体フィルムの熱的に成長させられたコントロールゲート509はチャンネル部分L2、ソースドレイン領域のシリコン表面から誘電体フィルム565aによって絶縁されている。その誘電体フィルムは誘電体567aと同じ材料で作られている。

pタイプの基板563の比抵抗は典型的には5から50オームセンチメータであり、p+のチャンネルドーピングされた560aの部分の濃度は典型的な $1 \times 10^{16} \text{ cm}^{-3}$ から $2 \times 10^{17} \text{ cm}^{-3}$ である。誘電体フィルム564aは典型的には20から40ナノメータの厚さで、誘電体フィルム567aは20から50ナノメータの厚さをもっている。フローティングゲート504aは通常N+にドープされたポリシリコンのフィルムであり、その厚さは25ナノメータ程度である（この厚さについては第4節で議論される）のであるが、25ナノメータをとるよりも低いか、または400ナノメータ程度の高さである。

コントロールゲート509は、ポリシリコンフィルムのN+に濃くドープされたものか、または低い抵抗の内部接続材料であり、例えば珪素または変化され難い金属である。

重要なことは、珪素または燐のイオン打ち込みにより形成されたN+ドレイン拡散領域502aの端部523aはフローティングゲート504aの端部522aに自己整合されたものであり、一方、同じようなイオン打ち込みのステップによって形成されるN+ソースのディフュージョン領域501aの端部521aは離れてはいるが、同じフローティングゲート504aの端部に自己整合させられており、それは、第4図には示されてはいないが、イオン打ち込み終了後でコントロールゲート509の形成の前に離される横壁スペーサを用いて形成されるのである。

拡散領域501a, 502aを形成するために用いられるインプラン特濃度は典型的には $1 \times 10^{15} \text{ cm}^{-2}$ から $1 \times 10^{16} \text{ cm}^{-2}$ のものである。

チャンネル部分L1とL2を形成するための基本となる大事なステップは第4b図から第4f図に図解されている。

第4b図の構造において、フローティングゲート504aと504bは薄いゲート酸化物564aの上に形成されたN+にドープされたポリシリコンの層により形成されており、このホトレジスト層590をマスクとして用いた反応性イオン異方エッチングによって形成される。

第4c図において、薄い保護フィルム566aが堆積によりまたは熱成長により形成され、続いて薄いスペーサ層57

0の堆積が行われる。

フィルム566aの主要目的は、その下に横たわる、例えは565aの層等をスペーサフィルムがエッチバックされるときにエッチされたり壊されたりすることを防ぐ目的で使用されているのである。

スペーサフィルムは注意深く時間を制御して異方性反応イオンエッチング段階によってエッチバックされる。

エッチバックの条件は意味のある切れ込みが生じないことであり、スペーサ材料と保護フィルム566aとの間のエッチ速度の比が20対1ぐらいのものである。スペーサ層570はドープされていないLPCVDのポリシリコンのコンホーマルフィルムであり、一方、保護フィルム566aは二酸化シリコン、シリコン窒化物であってもよい。これとは反対にスペーサ層570がLPCVDのシリコン酸化物のコンホーマルフィルムであり、一方、保護フィルム566aがLPCVDのシリコン窒化物、またはLPCVDのポリシリコンであってもよい。

保護フィルム566aの厚さはできるだけ薄い方が好ましく、典型的には10から30ナノメータぐらいが好ましく、続く珪素の打ち込みによりソースとドレインの拡散領域を形成するための浸透を許容するためにである。

コンホーマルなスペーサ層の厚さは横壁のスペーサの幅を決定し、そしてチャンネル部分L2の長さも決定される。典型的には400ナノメータのL2のために略600ナノメータの厚さのスペーサ層が使用される。

第4d図において、スペーサ592a, 593aと592b, 593bがフローティングゲート504aと504bの垂直な端面に沿って形成される。それぞれは、時間を計られたイオンエッチ段階の終了時にそれぞれ設けられる。

これらのスペーサは層570の厚さはフローティングゲートの壁、垂直な壁に近接する部分においてその平面部分より厚くなるという事実に基づいているのである。

かくして、注意深く時間を計った異方性のイオンエッチバック工程により570の平らな表面領域の部分のものはエッチで取られてしまうが、エッジに沿った部分については完全に除去されないで、その部分がスペーサを形成するのである。

狭い側壁スペーサをMOSトランジスタのゲートの側面に沿って形成する技術は当業界において公知であり、そしてそれは軽く僅かにドープされたドレイン（LDD）をMOSトランジスタの短いチャンネル製造工程において通常用いられている（例えば、1948年のアイイーディーエム（IEDM）の技術要約書の59頁にエス. メグロ（S. Meguro）等による「ハイシーモスIII技術（Hi-CMOS III Technology）」と題されるものに示されている）。

この発明において、スペーサはかなり広くとができる。そしてそれは、一方のエッジだけでも利用することができ、そしてそれは僅かにドープされたソース領域とかドレイン領域を規定するために用いられるものではなくて、一連のエンハンスメントトランジスタのチ

チャンネル部分L2を規定するため用いられるのである。

次の工程はマスキングの工程である。

ホトレジスト591a, 591b (第4d図) がスペーサ592a, 592bを保護し、スペーサ593a, 593bを露出するように使用される。後者はエッチによって除去される。好ましくは、湿式の化学エッチが用いられる (この化学エッチは保護フィルム566aをエッチしないように選ばなければならぬ。) そして、ホトレジストが剥がされる。

第4e図において、誘電体フィルム566aと565aを介して珪素の打ち込みがN+のソース拡散領域501a, 501bおよびN+のドレイン拡散領域502a, 502bを形成するために用いられる。

ドレイン側において、これらの拡散はフローティングゲートの端部522aと522bに自己整合される。

ソース側では、拡散はフローティングゲート505a, 505bの端部に自己整合されるが、スペーサ592aと592bの幅よりも離れ、横方向拡散は高温度プロセスに続いて行われる。

次にスペーサ592aと592bと保護フィルム566aは除去される (第4図f参照)。

好ましくは下部の層565aと504aを侵さないような式のエッチングが用いられることが好ましい。

誘電体フィルム567aは熱酸化またはLPCVDによる堆積によってフローティングゲートと基板の露出されている表面上に形成される。

導電層が続いて堆積される。そして、制御ゲート509aと509bが長い狭い筋のエッチングにより形成され、それは1つのアレイのメモリセルのワードラインの行を形成する。

残る処理工程は標準的なものである。

構造の表面は薄い表面保護膜568によって覆われる。そして、通常燐がドープされたガラスかボロン燐珪素ガラス (BPSG) が用いられる。

この保護層の形成は高温のアニール工程に引き続いて行われる。触媒バイアがエッチングによって構成される (これは、第4f図に示されていない)。これはソースとドレイン拡散領域への電気的接触を許容するためのものである。

金属の相互接続条帶569a, 569bは保護層568の表面に形成されて示されていないバイア開口を介してソースとドレイン拡散領域へのアクセスのために形成される。

トランジスタ500aの利点は以下のように要約することができる。

a) チャンネル長さ領域であるL1とL2の部分はマークの位置合わせずれに対しては影響を受けにくい。

したがって、これらを正確に制御することができる。

b) トランジスタ500aは横壁のスペーサを幅を制御することにより、L2を制御することができる。これにより、マスク合わせにより形成されるL2の部分よりもより短いチャンネルを制御しながら形成することができる。

これは、高い密度の分割チャンネルEepromとかフラッシュEepromトランジスタにおいて重要な配慮といえる。

前記第1の帯条の一部を除去するステップは隣接する第2の帯条間の空間を実質的に完全に伸びるマスク開口を介して行うエッチングによりなされるものであるが、これは公知のエッチング、例えばホトレジスト等を用いるもので、第1の帯条の一部を第2の帯条間でエッチングにより除去するものである。

前記第1の帯条の一部を除去するステップは前記第1の帯条を前記第2の隣接する帯条間の開口によってエンチングされるのであるが、その開口は前記間隔中の前記第2の帯条に沿って公知の誘電体のスペーサを形成することによって、フローティングゲートの第1の方向の両端が第2の帯条の上下方向に現れるようにしたものである。

I. b) 濃いドーピングしたチャンネルをドレイン接合の側にもつ分離チャンネル形のEepromトランジスタ

第10c図が本発明の第2の実施例に従う自己整合形でない分離チャンネルEepromトランジスタの断面図である。

第10a図と第10b図はこの製造するところの重要な工程を図示している。

トランジスタ1400はp形のシリコン基板1463 (これはp++基板上にエピタキシャル層成長によって構成されたp形のものであってもよい) から構成されている。

浅いN+のソース拡散領域1401とN+ドレイン拡散領域1402がI. a. で説明した実施例とは対照的にフローティングゲート1404を形成する前に形成されている。

ソースとドレイン拡散領域の間のチャンネル領域は2つの部分の領域に分けられている。

フローティングゲートの下に直接横たわっている部分L1 (1412) とコントロールゲート1409の下に直接横たわっている部分L2である。従来技術に対する改良点はドレイン拡散領域1402に隣接して濃くドープされた狭い領域1460をもつことに存する。

幅Lpの部分 (1413) とフィールドフェクトトランジスタのチャンネルが形成される表面の領域のドープ密度はp+が十分に高いということを条件として、この装置のプログラムと読み出し効率を制御するパラメータとなる。

典型的にはp基板1463はp形の $1 \times 10^{16} \text{ cm}^{-3}$ のドープ密度をもち、一方、1460のp+の領域は $1 \times 10^{17} \text{ cm}^{-3}$ と $1 \times 10^{18} \text{ cm}^{-3}$ の間のドーピング密度をもつ。

長さLpと領域1460のドーピング密度はプログラム電圧の条件下において、ドレイン接合における空乏領域の幅がLpの幅よりも小さくする。この条件が満足する限りにおいて、また、L1がLpよりも大きい限りにおいて現実のL1の大きさはこの装置において二次的な重要性をもつことになる。

パラメータLpがマスク合わせについてあまり影響を受

けないという限りにおいて、領域1460はドレインに自己整合させることができる。

そこで、L1はあまり重要でなくなり、Lpが制御されるパラメータになるのである。

パラメータLpをよりよく制御できる分離チャンネル形のEepromトランジスタ1400の新しい製造方法について説明する。

トランジスタ1400のメモリアレイの製造のためのこの新しい製造方法の主なステップは次のとおりである。

1. 第10a図に示す構造において、薄い酸化層1475は典型的には50ナノメータのシリコン酸化層であって、略100ナノメータ厚さのシリコン窒化物の層1474によって覆われている。そして、この層はシリコン酸化物の堆積によって形成された第2の層1473、その厚みは略100ナノメータ、により覆われている。

酸化物1475と窒化物1474は、例えば、メモリアレイの周縁領域の異方性分離領域を形成するときに使用される同じフィルムを用いてもよい。

2. ホトレジストのマスクP.R.1(1470)が、ホトレジストの開口部の端1471と1472間に幅をもち、長い平行な条の領域にソースとドレイン領域を規定するために用いられている。露出された酸化層1473は注意深く時間制御を行われたエッチステップによる湿式エッチングがなされる。そのエッチステップはホトレジスト1470を実質的に切り込む工程を含んでいる。

その切込みの限度、それは酸化物の端1476から1478の間の距離をLxとして測定されるものである、はパラメータLpの大きさを決定するために用いられる。

典型的なLxは300ナノメータから700ナノメータの間に選ばれる。再現可能なLxのための臨界的3つのパラメータはエッチ液(フッ化水素酸)の濃度と温度とエッチされるべき酸化物1473の密度(例えば、多孔性でないこと)である。

これらが十分によく制御されることにより、よく時間が計られたアンダーカットエッチングによりLxの幅がよく制御され、そして、それはホトレジストの開口部に沿う端部1471と1472に平行に走る。事実、500ナノメータ以下のLxの値のためにはホトレジスト層の長い狭い線の幅をコントロールするよりは横方向エッチングをコントロールすることによってより再現性の高いLxを得ることができる。

同じような横方向エッチングを利用する例は(これは大変に狭いガードリングを形成するために用いられるもので、同じ目的ではないのであるが)1979年のアイディエム(IEDM)の技術ダイジェストの49頁に表れるエス.キム(S. Kim)によって発表されたタイトルが「VLSI利用のための自己整合形のガードリングをもつ大変に小さいショットキーバリアダイオード」に示されている。

3. 横方向エッチング工程の完了時に第2の異方性エッチ

ングが行われる。その異方性エッチングは、同じホトレジストマスクP.R.1を用いて露出されたシリコン窒化物のフィルム1474の長い条を除去するものである。

P.R.1(1470)の端部1471, 1472は窒化物層のエッチされた条における各々の端1480, 1481を形成するために用いられる。

4. おおよそ $5 \times 10^{15} \text{ cm}^{-2}$ のイオン濃度をもつ砒素のイオン打ち込みが酸化フィルム1475を透過するに十分なエネルギーをもって行われ、N+にドープされた領域(1402, 1401)の長い帶条の表面をドープする。

ホトレジストのマスクP.R.1はこの工程におけるマスクとして使用されるが、窒化物1474も同様にインプラントマスクとして利用される。P.R.1はこの工程の後ではぎ取られる。

5. 打ち込みの障害のアニールと表面の酸化処理工程が引き行われ、200から300ナノメータの二酸化シリコン層1462がソースとドレインの拡散条の上に形成される。

この酸化における温度は領域1402, 1401におけるN+ドーパントの横方向の拡散を最小にするために、1000°C以下であるべきである。必要なならば、酸化フィルム1462をソースとドレイン領域だけでなく、フィールドを分離する領域にも成長させるために、さらに他のマスクステップで窒化層1474を除去することが可能である。

6. 第10b図において、第2のホトレジストマスクP.R.2(1482)が続くインプラントの工程において、基板のソース側(1401)を保護するために用いられる。

ボロンの打ち込みは、窒化物と1474と酸化物層1475を通過するに十分な比較的高いエネルギーで行われるが、表面の酸化物層1473と窒化物1474と酸化物1475を通過する程は高くはない。

これに対して、窒化物層1474は表面の酸化物1473の端部1478を用いて端1482に沿って最初にエッチングされる。

ボロンの打ち込みの量は $1 \times 10^{13} \text{ cm}^{-2}$ から $1 \times 10^{14} \text{ cm}^{-2}$ の領域にある。

濃くドープされたp+のドーピング領域の表面領域は大変に狭くて長い条として、その幅は表面の酸化物1478とN+拡散領域1402の端の間に延びており、ドレイン酢酸領域の条の長さだけ走っている。

厚い酸化物層1462がボロン打ち込み時にボロンがドレイン拡散条に入ることを防止していることに注意されたい。

これにより、ドレイン接合の容量を非常に減少させることができ、このことは、高速読み出しのために好ましいことである。

p+領域1460がこのプロセス中にドレイン領域1402に自動的に自己整合されることにも留意されたい。

7. 表面の酸化物1473、窒化物1474と薄い酸化物1475はエッチングによって除去される。このエッチングにより、ソースとドレイン拡散領域を保護していた酸化物層1462

の厚さをも減少させる。

このフィルムの厚さをこのエッティング工程の終了段階において略100ナノメータよりも薄くならないよう残すことが好ましい。

8. 残りの工程は、第10c図の構造に関連して理解されるべきである。

チャンネル領域を含み、長いソースとドレインの拡散条（典型的には15から40ナノメータの厚さの酸化物）であるゲート酸化物1462は、表面に成長させられる。

ポリシリコンの層（25から400ナノメータの厚さ）が堆積させられる。そして、N+にドープされ、フローティングゲート1404の連続的な細い条を形成するために、マスクされ、エッチされる。

マスクは、ドレイン拡散領域の条1402に平行で、1460のp+領域に重なるように位置合わせされる。

9. 第2の誘電体1466と1411がサブストレイトの表面とフローティングゲートの条の上にそれぞれ成長または堆積される。

この層は二酸化シリコンの層か、または二酸化シリコンの薄いフィルムとシリコン窒化物の結合によるものであって、それらの厚さは20から50ナノメータの間である。

10. 第2のポリシリコン層が堆積され、N+にドープされる（または、低い抵抗を与えるためには、珪化化合物）。

フローティングゲート条とソースドレイン条に直角に走る長いコントロールゲートの条1409を形成するためにマスクされてエッチされる。

各コントロールゲートの帯はフローティングゲートの帯に容量結合しており、各帯が互いに誘電体フィルム1411を介して重なっている領域において、容量結合している。

コントロールゲート1409は、フローティングゲート条によってカバーされていないチャンネル部分L2におけるチャンネルの導通を制御する。各コントロールゲートの帯は誘電体の分離フィルム（熱的に成長させられた酸化物）によってカバーされている。

11. コントロールゲートの帯をマスクとして利用することにより、誘電体1466, 1411の露出されている領域と最初のポリシリコンのフローティングゲートの帯がエッティングで除去される。その結果の構造は、コントロールゲートの長い帯または行である。各行はいくつかのフローティングゲート1404上に横たわり、そこにおいては、各々のフローティングゲートの外側の端が実質的にコントロールゲート上の幅を規定する端部に自己整合させられている。

これらの端部は酸化され、または、誘電体の堆積によりカバーされることにより、各フローティングゲートを完全に絶縁する。

隣接するセルの行間のフィールド領域またはソースと

ドレイン領域の間の隣接する帯はアクティブ装置領域に自己整合させられて、アイソプレーナ酸化による分離領域のようなスペースを必要としない。（もちろん、厚いアイソプレーナ酸化分離層により規定されるか、または、酸化物1462上のフィールド分離にたよるか、その酸化物というにはフィールド領域に成長させられたものであり、前述の工程の選択的な記述を参照されたい。これにより規定されるソースとドレインとチャンネル領域をもつトランジスタ1400を製造することも可能である。）

この実施例によるEpromセルは、先行するEepromセルに対して多くの利点をもっている。

a) コントロールゲート1409は、ソースとドレイン領域の上の比較的厚い酸化物層1462の上を走っている。

そのような厚い酸化物層はこれらのソースとドレインの領域がフローティングゲートが形成される前ではなく、後に形成されるいくつかの従来技術に示されているセルにおいては形成不可能なものであった。

この改良により、酸化物のブレイクダウンから保護し、コントロールゲートとドレイン間の寄生容量を減少させている。

b) パタメータLpの制御と領域1460における表面のp+ドーピング密度のコントロールは、前述したDSAの先行技術であるタナカのセルによって提供されるものよりもはるかに優れている。

c) フローティングゲートとドーピング間の位置ずれに対する装置の受ける影響度は、過去のセルによって経験されるものよりはるかに小さいものである。

d) チャンネル領域のp+の与えられた密度に対してドレイン接合の容量はすべての先行する装置のそれよりも小さい。なぜならば、p+領域1460は大変に狭く、ドレイン拡散領域のそばに閉じ込められて形成されているからである。

e) p+領域1460の濃度をエンハンスマント直列チャンネル領域L2の導通しきい値電圧を不当に上昇させることなく、非常に高いレベルにすることができる。（非常に高いレベルにすることによって、プログラムの効率を増大させることができる。）

このことは、このセルをEepromの部分として使用したフラッシュEepromの具体例において、極めて実質的に有効となる。

そのようなフラッシュEepromにおいては、フローティングゲート1404によってコントロールされる領域Lpのしきい値電圧は割合と高いものであり（初期電圧Vtは+5.0Vぐらいの高さであって、これは供給電圧またはそれ以上に相当するものである。）、これはセルを低いしきい値電圧に消去することによって克服される。

Eeprom装置において、初期電圧Vtはプログラムされていない状態において、読み出しの期間中には、コントロールゲートの電圧よりも高くてはいけないのであり、この要請はp+ドーピングの密度をどのくらい高くできる

かという上限を規定している。

他の1460のp+ドーピング密度の大きさを規定する限界は、プログラミングのために必要な最低ドレイン電圧によって確立される。

このドレイン接合のアバランシュブレイクダウン電圧は、この最小プログラム電圧と少なくとも同じくらいの大きさでなければならない。

## II. アイソプレーナフィールド分離を用いた自己整合形チャンネル分離フラッシュEEPROMセル

第5a図は、本発明の具体例にしたがう2×2のメモリセルアレイ、つまり4個のフラッシュEEPROMトランジスタ600a, 600b, 600cと600dを含む装置の平面図を示している。

第5b図は、第5a図の線AAに沿う断面図を示している。

線BBに沿う第2の断面図は第4図に示されているEEPROMトランジスタ500aを断面したことになっている。

第5a図に示されるトランジスタ600aは分離チャンネルEEPROMトランジスタであって、それには消去ゲート530, 535が付け加えられている。その消去ゲートはフローティングゲート504aの端部、532a, 562aに重ねられている。

トランジスタ600aはソース拡散領域501a、ドレイン拡散領域502aとコントロールゲート509をもつ分離チャンネルのEEPROMトランジスタとしてプログラムされている。

フローティングゲート504aとチャンネル部分L1とL2はI.a.節で述べた分離チャンネル方式のEEPROMトランジスタ500aまたはI.b.節で述べた分離チャンネルEEPROMトランジスタ1400と同じような方法にしたがって形成されている。

しかしながら、他の分離チャンネルのEEPROM構造として利用することができる。

トランジスタチャンネルの幅Wは厚いフィールド酸化部の562の端部505, 505aによって規定される。

トランジスタ600aはフローティングゲート504aから消去ゲート530, 535への電流のトンネリングによって消去される。

消去ゲートが重ねて設けられているフローティングゲートの表面と横壁に沿って設けられているトランジスタ誘電体531a, 561aの横切って流れるトンネル電流によって消去される。

トンネル誘電体フィルム531aと561aはフローティングゲートを含み、N+にドープされて纖維化された多結晶シリコンの熱酸化により成長された二酸化シリコン層により形成される。このことは、当業界ではよく知られている。(例として、エイチ.エイ.アール.ウェグナー(H. A. R. Wegener)によってIEEEの国際電子装置会議1984年の12月の第480頁に「テクスチャードーポリクリスタルシリコンのフローティングゲート記憶素子の耐用年月モデル」という論文に示されている。適当に纖維化されドープされた多結晶シリコンを適当な酸化条件の下に生

成させたときにそのようなフィルムは単結晶シリコン(そのようなトンネル誘電体フィルムはサマチサ アンド キネット(Samachisa and Kyenett)の装置に用いられている)よりも数倍厚くてもそのようなトンネル電子の誘電体を数倍に増大させることができる。

例えば、纖維化されたポリシリコンで、N+にドープされた40ナノメータの厚さに成長させられた誘電体トンネル酸化物は、理想的な電圧の印加された状況において、N+のドープの単結晶シリコンの10ナノメータの厚さのトンネル誘電体の電流密度と略同じ電流密度が得られている。

この高効率のトンネリングの発生する機構は、ポリシリコンの粒子境界における鋭いギザギザ、特にそのような纖維構造が領域的にそのようなギザギザの密度を増大させるということが、高いトンネル効率を与えるものと信じられている。

よく行われている技術は、まず始めにポリシリコンの表面を高い温度により酸化してギザギザを引き立てて、それから酸化物を引き離してトンネル酸化物を低い温度によって再成長させることである。

そのようなゴツゴツさせられた部分をもつ酸化物フィルムは部分的に独立して4から5倍の印加された電界の4倍から5倍の部分的な増幅をし、その結果として部分的なトンネルインジェクタとなる。

より厚いトンネル誘電体フィルムの利点は、それらが均一に成長させられ易く、かつ、欠点のない層が得られることである。

さらに、厚いトンネル誘電体(40ナノメータ程度)でトンネル中に発生する電界は薄いトンネル誘電体(10ナノメータ程度)中で発生するものの25%程度であるが、ただし、これは同じ電圧バイアス条件が与えられたときである。

このことは、消去サイクルにおいてより高い信頼性とより強い耐性が得られることに帰結する。

これらの利用により、本発明のすべてのフラッシュEEPROMの実施例は比較的厚いトンネル誘電体を介して多結晶一多結晶状態に依存した消去を行うのである。

第5a図、第5b図の実施例において、フローティングゲート504aは濃くN+ドープされた厚さ25から400ナノメータの間のポリシリコンの層上に形成される。

消去ゲート530, 535は厚さが50から300ナノメータのN+にドープされたポリシリコンの第2の層上に層の中に形成される。

そして、コントロールゲート509は、厚さが200から50ナノメータの間にある第3の導電層中に形成される。

この第3の導電層は、N+にドープされたポリシリコンか、または、チッ化珪素または金属メタルであり得る。

消去ゲートは比較的薄い層に形成することができる。なぜならば、比較的高いシート抵抗(例えは100オーム

/スクエア) は、トンネル消去の間中にはこのゲートの中にはほとんど電流が流れないのであるから、耐えられるのである。

製造段階は、消去ゲート530と535をコントロール509のために使用されるのと同じ導電層中に構成することによって、いくぶん簡略化することができる。

しかしながら、コントロールの端と消去ゲートの端までの間隔 $Z$  (すなわち、セルサイズ) はコントロール消去ゲートが2つの異なった導電層に亘りに誘電体フィルム567aによって絶縁されて形成されているケースに比べてかなり大きくなる。

事実、第5a図の600aの3層の構造においては、コントロールゲート509を消去ゲート530と535の上に僅かに重ねるようにすることさえも可能である (すなわち、スペーシング $Z$ は零またはマイナスになるのである)。

トランジスタ600aは、フィールドを分離するための酸化物562 (第5b図参照) をもち、その厚さは200から1000ナノメータである。ゲート酸化物564aを保護するチャンネル部分L1 (512a) は、熱的に成長された二酸化シリコンであり、その厚さは15から40ナノメータである。

誘電体フィルム567a、それはコントロールゲート509に強く容量的に結合させられており、フローティングゲート504aが成長または堆積されている。それは二酸化珪素または薄い二酸化珪素フィルムと酸化されたシリコン窒化物の組合せによる厚さ20から50ナノメータのものである。

この誘電体はゲート酸化物を保護するチャンネル部分L2 (520a) およびソースとドレイン拡散領域の上の絶縁部565a (第4a図参照) として機能する。

消去誘電体531aと561aは熱的に成長させられた二酸化シリコンまたはその他の積層された誘電体であって、それはシリコン窒化物のように十分な消去導通特性をもっている。その厚さは30から60ナノメータである。

重要な1つのことは、531aと561aの結合領域からなる各々のセルの消去に寄与するこのトンネル誘電体領域がエッジ532a, 562a、つまりこれはフローティングゲート504aのエッジと消去ゲート530, 535間のマスクの不一致の影響を受けないということである。 (各々の消去ゲート、例えば、535は2つの隣接するセル、この場合においては600aと600cによって分割されて所有される)。

そのような合わせのずれはフローティングゲートの一端のトンネル誘電体の領域の減少に原因するものであるが、しかし、フローティングゲートの他端におけるトンネリングに利用される領域における同量の増加として表れる。

この特徴によりセルの構造をトンネル誘電体の非常に狭いトンネル誘電体の領域をもって実現することができる。

先行技術として示したマスオカとクオの3層のフラッシュEEpromセルと対比すると、それらはマスク合わせの

それに対して非常に敏感であり、その結果、それらにおいてはトンネル消去においては理想的な領域、より以上に大きい領域、を非常な極端な合わせずれの場合を想定して取っておかなければならなくなるのである。

メモリセル600, 600b, 600cと600dを消去するために必要な電圧の条件は次のとおりである。

$V_{\text{ERASE}}$  (すべての消去ゲート530, 535, 536において) は、15Vから25V, 100ミリセコンドから10セコンドである (パルスの持続時間は $V_{\text{ERASE}}$ に強く依存して決まる)、 $V_{\text{CC}}=0V$ ,  $V_{\text{BB}}=0V$ ,  $V_{\text{D}}$ と $V_{\text{S}}$ は0Vに保たれるか、または高い電圧では5Vから10Vの間の電圧を選ぶ。それは、消去期間中において、領域563のようなところの誘電体フィルム565aが経験する電圧を少なくするためである。

そこでは、ゲート530はドレイン拡散領域502の上をまたいでいる (第5a図参照)。

### III. フィールドプレート分離を用いた自己整合形分離チャンネルフラッシュEEprom素子

本発明の他の実施例にしたがう $2 \times 2$ のフラッシュEEpromセルは第6a図に平面図として、そしてAAとCCの断面図として第6b図と第7c図にそれぞれ示されている。

BBの断面図は第4a図の分離チャンネルEepromトランジスタと実質的に同じである。

分離チャンネル形のフラッシュEEpromトランジスタ700aは3つの導電層 (フローティングゲート704, 消去ゲート730, 735とコントロールゲート709) を使う。

そしてそれは第2節でフラッシュEEpromトランジスタ600a、つまり第5a図、第5b図に示されているものに関連して説明した順序でそれらの層は形成されている。

トランジスタ700aの主たる特徴は消去ゲート730, 735, 736がトンネル消去に使われるだけではなく、アクティブトランジスタ領域の外に形成されたフィールドトランジスタの分離のためのゲートのスイッチオフに使用されることである。

かくして、セル600a (第5b図) の厚いアイソプレーナ分離酸化物562は不要となる。

アイソプレーナ分離酸化物562は、メモリセル700a, 700b, 700cと700dのアレイの中でより薄い酸化物762 (第6b, 7c図) によって置き換えられており、フィールドプレート730, 735, 736によって覆われており、それは消去期間を除いて、常に0Vに保たれている。

メモリセルのアレイの内側の厚いアイソプレーナ酸化物を除去することは、多くの利点がある。 (このアイソプレーナ酸化物は周辺の論理トランジスタとの間に分離のために残されることもある。)

1. 長い熱のアイソプレーナ酸化サイクルに起因する二酸化シリコンの環境領域における表面のストレスはアレイの内部において消去されるその結果、ソースとドレイン接合における漏れが少なくなり、ゲート酸化層の良い品質が得られる。

2. 与えられたセルの大きさの範囲内で異方性の酸化層を

除去することにより、有効なチャンネル幅 $W_1$ が、フローティングゲート704の下で、フローティングゲートの2つの端732aから762aの上を延びるようにすることができる。比較すると、トランジスタ600a（第5b図参照）の有効チャンネル幅 $W$ はアイソプレーナ酸化の端505によって決定されるものであるからして、実質的に小さくなる。

この差異に起因して、セル700aにより高い読み出し信号を得ることができ、より狭くすることができ、セルをより小さくすることができる。

3. トンネル消去の効率のための容量結合は、フローティングゲートとシリコン基板763の接合が最も大きい。トランジスタ700aにおいて全体の底面の面積、フローティングゲートの面積は、薄い誘電体764を通じて基板763に密に結合している。

これと対照的に、トランジスタ600a（第5b図参照）のフローティングゲート500aの多くの底面は、厚いフィールド酸化物562aの横に横たわっているために、基板563には強く容量結合させられていない。

4. コントロールゲート709のその端部744から774の間の幅は、第6c図に示す直列エンハンスマント形のチャンネル部分L2の幅 $W_2$ を規定している。

このことは、異方性酸化物の端に重なるコントロールゲートの要請を除去するということに起因してすべてのセルの幅を小さくすることができる。セル700aを製造するにあたって、前もって注意すべき点は、ゲート704aのエッジ732a、消去ゲート730の端部784、コントロールゲート709の端部744を規定するマスク層との間の合わせずれが分離チャンネルL1とL2に並列にコントロールゲート709の下に狭い規制端のトランジスタが形成されるような状況が許容されてはならないということである。

しかしながら、セル600aについてはその消去ゲート730、736とコントロールゲート709は2つの分離された導体層に形成されているので、その2つの層はお互いに誘電体絶縁フィルム767（第6b図参照）により分離されているから、エッジ784とエッジ744間の空間的な分離Zの大きさについては何らの要請もない。

事実、2つのエッジは互いに重なってもよいし、大きすぎてもよいし、位置合わせずれが起こってもよい。すなわち、Zは零または負であり得るのである。誘電体絶縁体767はゲート誘電体766（第6c図参照）の一部を形成し、チャンネル部分12の上にある。

メモリアレイにおいてソース拡散領域701とドレイン拡散領域702は長い帯条に形成されてもよい。もし、トランジスタ500aがEepromトランジスタとして使用されるのであれば、ソース拡散領域の端部721は前述したように横壁のスペーサ（図示されていない）に自己整合され一方、ドレイン拡散端723はフローティングゲート704aの端部722に自己整合させられる。フローティングゲート704a、704c近傍に隣接する領域において、ソースとド

レインの拡散端（721x、723x、第6a図参照）は互いにもぐり込むことが防がなければならない。

これは例えば、最初にポリシリコンの長い連続的な帯の一部としてフローティングゲート704aと704cを形成し、それから、長い連続的な横壁のスペーサをもつこの帯を用いてイオン打ち込みにより長い拡散の帯701、702を形成する。それからスペーサの帯を除去し、そして単に、ポリシリコンの長い連続した帯を、分離されたフローティングゲート704a、704cを形成するために、長い連続的なポリシリコンの帯をエッジ732a、762aに沿ってエッチングする。

前述したフラッシュEEPROMの実施例についても、この実施例をEEPROMセル1400（第10c図参照）と関連して用いることができる。

IV. フローティングゲートの垂直の端に閉じ込められた消去をもつ自己整合形のチャンネル分離形のフラッシュEEPROMセル

本発明による自己整合形のチャンネル分離形のフラッシュEEPROMのさらに他の実施例においては、第II節、第III節において記述した実施例のセル600aと700aのそれぞれよりも小さい領域をもつものである。

この第3の具体例において、フローティングゲートと消去ゲート間のトンネル消去の領域は、各フローティングゲートの2つのエッジに沿う垂直な横壁の表面に実質的に閉じ込められて形成されている。

この実施例のセル800aがセル700aとセル800a、800b、800cと800dの2×2アレイセルから、どのように異なるかということの最もよい理解のために、第7a図は平面図を示しており、第7b図は第7a図の方向AAに沿った断面図であり、このAAに沿った断面図はセル700a、700cの第6b図に対応するものである。

セル800aは、N+に強くドープされたポリシリコン第1の層の中にフローティングゲート804aをもっている。このゲートはチャンネル部分L1（第7a図参照）のトランジスタの導通を制御するものであり、それは、酸化絶縁フィルム864のゲートを通して行われる。

コントロールゲート809は、第2の導通層中に形成されている。そして、それは誘電体フィルム867によりフローティングゲートから絶縁されている。そして、その誘電体フィルム867は熱的に成長させられた酸化物であるか、または、薄い二酸化珪素とシリコン窒化フィルムの結合であり得る。

コントロールゲート809の端874と844は、それぞれフローティングゲート804のエッジ862a、832aに自己整合エッチングする領域を規定するマスクとして使用される。

消去ゲート830、835は第3の導電層内に形成され、そして、それらはフローティングゲート804aの端832a、862aに重ねられる。各消去ゲートは、例えば830は2つの隣接するセル（800a、800cのような）によって分けもたれている。

消去ゲートは、消去ゲート830, 835, 836の積層に先立って、成長または積層される誘電体絶縁層897によってコントロールゲート809から絶縁されている。

トンネル消去誘電体831a, 861aはフローティングゲート804aの垂直な端面832a, 862aの表面に閉じ込められて形成されている。

構造800における導電および絶縁層の厚さは700aにおけるそれと略同様である。

しかしながら、消去ゲートはコントロールゲートよりも前ではなく、後に作られるのであるから、製造工程はかなり異なっている。

特に（第7a図、第7b図を参照されたい）、

1. フローティングゲート804a, 804cはゲート酸化物864の表面に長い連続的な狭い帯として形成されている。そのような各帯の幅は、L1に、ドレイン拡散領域にフローティングゲートが重なる領域の程度をえたものである。
2. 誘電体層867が形成され、それから第2の導電層（ポリシリコンまたはシリコンをN+にドープしたもの）が積層される。
3. コントロールゲート809がフローティングゲートの帯の方向と直角な方向に狭い帯として規定される。

この帯は、端部844, 874に沿ってエッチされ、そして比較的厚い誘電体層897によって絶縁される。

4. 端部844, 874（または、コントロールゲート帯809の両端間に形成された絶縁スペーサ899の端）は、誘電体867をエッチするのに使用される。それから、自己整合によりフローティングゲート帯の端832aと862aに垂直な端をエッチする。その結果、これらの垂直な壁に沿って露出されたポリシリコンの壁をもつ絶縁されたフローティングゲートが形成される。

5. トンネル誘電体フィルム831a, 861aがこれらの露出された表面の熱酸化によって形成される。

6. 第3番目の導電層が積層される。それから消去ゲート830がコントロールゲートの帯に隣接して平行にその間を走るように形成される。

これらの消去ゲートは、メモリアレイの隣接する領域を電気的に絶縁するための領域絶縁プレートとして機能する。

フラッシュEEPROMトランジスタ800aは、この発明による分離チャンネルEEPROMトランジスタ（トランジスタ500aと1400a）の両側のいずれかに関連して形成される。

例えば、フラッシュEEPROMトランジスタ800aの列は、分離チャンネルEEPROMトランジスタ1400（第10c図に示す）の製造工程にわずかの工程を付加することによって形成される。

次のとおりである。

分離チャンネル形EEPROMトランジスタ1400を製造する工程の1から10の工程は第I. b. 節に関連して記述された製造方法の1から10と同じである。

ステップ11, 12, 13は、第4節において、分離チャンネ

ルフラッシュEEPROMトランジスタ800aの製造に関連して述べられた製造工程4, 5, 6と同じである。

セル800は非常に小さいトンネル消去の領域をもつものとして形成され、そして、そのコントロールは極めて容易である（それはマスクの大きさによって規定されるのではなく、むしろフローティングゲートを形成する堆積層の厚さによって決められる）。

この理由により、このセルは、この発明の実施例においてもっとも高くスケールされるものといえる。

V. 埋め込まれた消去電極をもつ自己整合分離チャンネルフラッシュEEPROMセル

本発明の第4の実施例によるフラッシュEEPROMセル900a, 900b, 900cと900dの2×2のアレイの平面図を第8a図に示し、2つの線AAおよびDDに示す切断図を第8b図および第8c図にそれぞれ示す。

第8a図の断面BBは第4a図の分離チャンネルEEPROM構造500aの構造と同じ断面となっている。

トランジスタ900aは、分離チャンネルフラッシュEEPROMトランジスタであって、EEPROMトランジスタ500aのように、自己整合により、またはEEPROMトランジスタ1400のように自己整合でない方法によって形成されたチャンネル部分L1とL2をもっている。

消去ゲート930は狭い導電性の帯であって、フローティングゲート904aを下にしてコントロールゲート909aを上にしてサンドイッチされた状態で形成されている。

消去ゲート930はフローティングゲートの端932a, 962aから離れて位置させられている。これらの端部はそのためトンネル消去の役割を果たさない。トンネル消去はフローティングゲート904aに重なる消去ゲート930の領域に閉じ込められたトンネル誘電体931によって行われる。

消去ゲート930は直列エンハンスマントチャンネル部分L2に幅W<sub>o</sub>だけ重なっている。

読みまたはプログラミングの期間において、消去ゲート930はOVに保たれている。その結果、幅W<sub>o</sub>のチャンネル部分は、読みまたはプログラム電流に何ら寄与しない。

チャンネル部分L1で導通に寄与する部分は、W<sub>o</sub>とW<sub>o</sub>の部分であって、その部分はコントロールゲート909によって直接制御される。

チャンネル部分L1は、しかしながら、すべての3つの幅W<sub>o</sub>, W<sub>o</sub>とW<sub>o</sub>からの導通の寄与を見ることになる。

フローティングゲート904aの端932a, 962aはそれぞれコントロールゲート909のエッジ944, 947に自己整合でエッチングができる。

これにより、コントロールゲートまたはフローティングゲート（第8b図）によって保護されていないフィールド領域にpタイプのドーパントを打ち込みすることによりチャンネルストップ領域の分離998を形成することが許容される。

セル900aの1つの特徴は、これらのセルの帶条を形成

する導電層の等方性エッティングにより、コントロールされた切込みを利用することにより、消去ゲートの帯930, 936を非常に狭くすることができることがある。

その結果、トンネル消去の非常に狭い領域として表され、それはマスク合わせのずれには敏感でない。加うるに、チャンネル幅 $W_p$ と $W_n$ はマスクの不整合に対して同様に敏感ではない。

#### VI. 新しい消去機構をもつ端が調整されたフラッシュEEprom

本発明の具体例であるフラッシュEEpromの実施例600a, 700a, 800a, 900aはポリシリコンのフローティングゲートの纖維化された表面に形成された比較的厚い誘電体酸化物を通して行われるトンネル消去が利用されている。他のものは小さいギザギザがあること、デコボコしていること、略30ナノメータの直径の曲率をもっていることが（この場合においては、フローティングゲートであるが）、インジェクタサーフェイスの電界を4ないし5倍上昇させることを前提として、比較的厚いトンネル誘電体フィルム（30から70ナノメータ）をとおして、効率のよいトンネル導電が行われることを仮定している。したがって、従前の技術においては、ポリシリコン表面の高温酸化を用いてポリシリコンの表面がギザギザになるように成形する技術等が用いられていた。

しかしながら、そのような工程は再現可能であるが、本来体験的なものであって、装置が割高となり、そして、そのこと自体はよりよく理解されてはいない。

本発明によって新しいアプローチが開示されるのであるが、それは非常に再現性がよく、電気的な電界によるトンネル消去を増強したものであるが、それらは従来のギザギザとかデコボコによる方法よりもより効果的であり、EEpromとかフラッシュEEprom装置により簡単に適用できるものである。

このアプローチにおいて、フローティングゲートの層は典型的には25から200ナノメータのように非常に薄い層で積層される。

これはすべての従来技術のEeprom, EEpromとかフラッシュEEpromのフローティングゲートより薄いものであり、従来のものは、少なくとも200ナノメータ程度の厚さのポリシリコンの層を用いていたのであり、通常は350から450ナノメータであった。

従来技術において、ポリシリコンの厚さが200ナノメータより厚く選ばれていた主たる理由は、低いシート抵抗とポリ酸化物のよい品質は厚いポリシリコンによって提供されるということに基づいている。

ある従来の装置においては、フローティングゲートは打ち込み用のマスクとしても用いられるのであるから、それは打ち込み用イオンの浸透を妨げるために十分な厚さをもっていなければならなかった。

このようにして、チャンネル分離のEepromの具体例500a（第4a図参照）において、スペーサの形成（第4b図か

ら第4f図参照）は、フローティングゲート504aが100ナノメータまたはそれ以下の厚さである場合は、容易にできないことであった。4

しかしながら、Eepromトランジスタ1400（第10c図参照）とフラッシュEEpromトランジスタ600a（第5a図参照）, 700a（第6a図参照）, 800a（第7a図参照）と900a（第8a図参照）は、クベックの先行技術のトランジスタ200b（第1図参照）と同様に消去の効率において、有意義な改善を達成するために、100ナノメータの厚さのフローティングゲートを構成していた。

そのような薄いポリシリコンの層を形成しようとする理由は、そのような薄い層のフローティングゲートのエッジは酸化の過程を通じて調整することより、極端に鋭いチップの端にできることであった。

このような尖った端部の曲率は極端に小さくすることができ、それはトンネル用の誘電体成長の厚さと同様にポリシリコンのフィルムの厚さによって規定されるものである。

したがって、これらの鋭い先端によるトンネル消去はもはや表面の粗さに従属することなく、そのトンガリそれ自体によって規定されている。

この変形例の図解のために、フラッシュEEpromトランジスタ800a（第7a図参照）を2つの異なった実施例として考慮し、比較的に厚いフローティングゲート（トランジスタ800aは第7b図と第12a図に示されている）と同じトランジスタを非常に薄いフローティングゲートをもつものとして改良したものであり（第11b図にトランジスタ800Mとして示してある）。第12a図の断面図において（これは第7a図のAA方向の断面に相当する）フローティングゲート804aは略300ナノメータの厚みである。その垂直方向の端862aと832aは表面に小さなギザギザが設けられていることが示されている。

これらのギザギザはトンネル消去の期間に電子を注入するものとして作用する（トンネル誘電体層861a, 831aを横断する矢印として方向が示されている）。

注入された電流は消去ゲート835, 830はにより集められるのであるが、その消去ゲート835と830は862aと832aの垂直方向のエッジに重ね合わされている。

これに対照して、改良されたトランジスタ800Mの断面図は第11b図に示されており（これは第7a図のAA断面と同じ断面である）、トランジスタが100ナノメータまたはそれ以下の厚さのフローティングゲート804Mをもっていることを示している。

誘電体層864と867は、トランジスタ800aのコントロールゲート809と略同様である。

トランジスタ誘電体層861M, 831Mを形成するためのフローティングゲート804Mの薄い垂直端の酸化の過程において、薄いフローティングゲートの上および下面是その露出された端において酸化される。この結果、非常に尖ったチップ870lと870rが形成される。

このチップは非常に優れた効果的な電子注入部として機能する（トンネル誘電体861M, 831Mを横切る矢印として示されている）。

注入された電子は、トランジスタ800aの消去ゲート835, 830、それらは非常にこれらの鋭いチップに重ね合わされている消去ゲートによって集められる。

トランジスタ800Mの非常に薄いフローティングゲートに固有に授けられた特性で、極めて効果的で高い再現性をもつという特徴と離れて、トランジスタ800aを含むすべての実施例における対応する容量よりも非常に少ない容量よりもそのフローティングゲートとそのチップ間にもつという付加的な利点をもつものである。

他の2つの特徴は特筆すべきである。

まず、第1に、ポリシリコン804Mとゲート誘電体864を通じてN+のドーパントの侵入を防止するためにこの非常に薄いフローティングゲートはあまりドープされるべきではない。

フローティングゲート804Mはけっして電流導体として利用されるものではないから、そのシート抵抗は100から10000オーム／スクエアが好適である。

第2に、コントロールゲート809Mと同様に基板860またはソースまたはドレイン拡散領域（第11b図には示されていない）からフローティングゲートの鋭いチップは十分に離れるか、または絶縁されていなくてはならない。

鋭いチップのメカニズムは非常に高い機能をもっているから、希望しない部分的を消去してしまうということが、装置をプログラミングする（すなわちプログラムディスターバンスの距離条件）において存在する電圧条件下において起こる。

このことは非常に薄いフローティングゲートを端部の消去のために使用するいかなるフローティングゲートトランジスタの現実的な幾何学的配列の際に考慮されるべき重大な事項である。

しかしながら、薄いフローティングゲート層は比較的鋭いチップ端を持つ端の酸化を達成するための直線的なアプローチ、比較的厚いフローティングゲート層の中にも鋭いチップ端を形成することが可能である。

例えば、第12c図において比較的厚い層がフローティングゲート804を形成しており、それは引込み型のエッティングがなされている。酸化の後に鋭いチップ781が上側の端に形成され、高い電界のトンネリング861を、トンネル消去誘電体831の上に形成された消去ゲート832に与える。

第12d図の装置において、消去ゲートはフローティングゲートの前に堆積される。消去外831はその底面に近接して引込みキャビティを作り出すようにエッティングされる。

トンネル消去の誘電体層831はそれから形成され、それに引き続いてフローティングゲート804を形成する積

層が行われる。

フローティングゲート804は鋭いチップ870が形成されるべき狭い引込み形の空洞を埋める。これにより、高い電界トンネル861が形成される。

第12d図に示す装置は、消去ゲートの表面にギザギザをもつことになるのに対し、本発明において記述された他の装置はフローティングゲートの表面にギザギザをもつという点について注目されたい。

#### VII. フラッシュEEPROMメモリアレイの構成

本発明によるフラッシュEEPROMセルは、いろいろの異なったアレイ構造において、高密度のメモリアレイとして集積化される。

第11a図は第1の構成に示しており、これはEEPROMアレイの業界において通常よく用いられるものである。

第11a図の3×2のアレイは2行3列のフラッシュEEPROMトランジスタを示している。

トランジスタT10, T11, T12、すなわち第1行のトランジスタは共通のコントロールゲート（ワード線）と共にソースSをもっている。各ローのトランジスタは、それ自身のドレインDをもっており、それは、列のビットラインに接続されていて、それらと同じ行の他のすべてのトランジスタのドレインと共有されている。

すべてのトランジスタのフローティングゲートはすべてのトランジスタのドレインに近接して設けられており、それらのソースからは離れている。消去線はビットライン方向に走っていることが示されている（ワード線の方向に走らせることもできる）。各消去線はトランジスタのフローティングゲートの左側へ、そして消去線の右側から結合されている（消去誘電体層を介して）。

異なった動作モードにおける電圧条件についてテーブル1（第13a図）に示してある。

すなわち、選択されたセルだけではなく、選択されないセルについても同じ（ワードライン）を共有するもの、または、同じ列（ビットライン）を共有するものについて示されている。

アレイ中のすべてのセルのブロック消去の期間において、すべての消去ラインはハイレベルにされる。

そのセクタの中の一対の消去ゲートのV<sub>ERASE</sub>をハイに保ち、他の消去ゲートをOVに保つことによって、アレイの僅かな位置セクタを消去することが可能である。

第11a図に示したアレイよりも、よりよい実装密度を上げることができるフラッシュEEPROMアレイの第2の構成は下層接地アレイとして知られている（このアレイの構造についてより詳しい理解をするためには、ハラリ（Herari）の米国特許第4,328,565号を参照されたい）。

そのようなセルのアレイの平面的な分布は第5a図、第6a図、第7a図と第8a図に示されている。

第5a図のアレイに対する2×2の仮装接地メモリアレイの略図的な表示は第11b図に示されている。

仮装接地アレイにおいては、ソースとドレインの領域は交換可能に用いられる。

例えば、ディフージョン502はトランジスタ600aのドレインとして利用することができるし、トランジスタ600bのソースとして利用することもできる。

ここにおける、言葉「仮装接地」は、接地の供給が現実の接続によらないで、ソースに供給されるべきグラウンドがハードワイヤによるものではなくで、デコードされるべきものであるから、このようの言葉が生じたのである。

このデコーディングがソースをグラウンドラインまたはドレインと交換に用いることを許容している。

仮装接地アレイにおける動作条件はテーブルII（第13b図）に示されている。

これらは、実質的に標準的な構造と同じである。異なる点は、選択されなかつたセルのすべてのソースとドレインのコラムは、プログラムの期間において自己によるプログラミングの弊害を防止するために浮いた状態に保たれていることである。読み出しの期間においてすべての列は、低い電圧（略1.5V）に引き上げられる。そして、選択されたセルだけが、そのソース拡散領域をもじ接地電圧に近いところまで引き下げられ、その電流はセンスできるようにするのである。

アレイはブロック単位において消去できる。または、すべての行において消去すべき電圧を消去線に対応してデコードすることによって消去することができる。

ここで説明された実施例は本発明の好適な実施例として説明されたものであり、当業界はこれの種々の変形例について理解できるであろう。

特に、分離チャンネル形のフラッシュ EEPROM装置600a, 700a, 800aと900aは、第4b図から第4f図に略図示された一方のサイドスペーサにしたがって構成されたチャンネル部分L1, L2をもつトランジスタ500aの分離形チャンネル EEPROMに関連して形成されるか、またはEEPROMトランジスタ1400、または他の自己整合プロセス技術または自己整合方法によらない先行技術によって形成されたEEPROMトランジスタのようなものによって容易に形成できるものである。

したがって、本発明は、添付の特許請求の範囲の全内容にしたがって保護が与えられるべきである。

#### 【図面の簡単な説明】

第1図はクペック等による三重ポリシリコンEEPROMセルの先行技術を示す断面図である。

第2図は米国特許第4,531,203（マスオカ（Masuoka））の三重ポリシリコンチャンネル分離形のフラッシュEEPROMセルの先行技術の分布図である。

第3図は先行技術のセルであるドレインジャンクションに近接する非常に高いドープをなされたチャンネルをもつ積み重ね形ゲートのEEPROM（タナカ（Tanaka））に関する断面図である。

第4a図は本発明による分離チャンネルEEPROMの断面図である。

第4b図から第4f図は第4図に示したセルの製造過程における種々の工程の断面図である。

第5a図はこの発明の実施例に従う三層の構造のフラッシュEEPROMセルの2×2個のアレイ、アレイ中の2×2の平面図である。

第5b図は第5a図の構造の切断線AAに沿って切断して示した断面図である。

第6a図は本発明の第2の実施例に従う三層の構造のフラッシュEEPROMセルの2×2のアレイの分布図であって、ここにおいて消去ゲートが分離のためのフィールドプレートを提供している。

第6b図は第6a図の構造の切断線AAに沿って示した断面図である。

第6c図は第6a図の構造の切断線CCに沿って切断して示した図である。

第7a図は本発明の第3の実施例に従う構造の三層に形成された構造のフラッシュEEPROMセルの2×2アレイの平面図であって、ここにおいてトンネル形の消去誘電体層がフローティングゲートの2つの端面の表面が垂直領域に限定されて設けられている。

第7b図は第7a図の構造のAA線に沿って切断して示した図である。

第8a図は本発明の第4の実施例に従う三層構造に作られたフラッシュEEPROMセルの2×2個のアレイの平面図である。

第8b図は第8a図の構造のAA切断線に沿って切断して示した図である。

第8c図は第8a図の構造のDD切断線に沿って切断して示した図である。

第9図は本発明のフラッシュEEPROMセルのフローティングゲートに関する結合キャパシタンスを略図的に示した図である。

第10a, 10bおよび10c図は本発明による実施例装置の製造段階における重要なステップにおける切断断面図である。

第11aと11b図は本発明によるフラッシュEEPROMの実施例の2つのメモリ素子を略図的に示した図である。

第12aと12b図はそれぞれフラッシュEEPROMトランジスタの断面図を示すものであり、第12a図はアスペリティインジェクションの消去構造を、第12b図シャープチップインジェクションの消去構造を示している。

第12cと12d図はフラッシュEEPROMトランジスタの部分断面図であり、高電界の電子注入をする直接エッチングによるフローティングゲートのシャープチップトエッジの構成を示している。

第13a図には表1が示されており、この表は第11a図のアレイのすべての動作モードの電圧条件を示している。

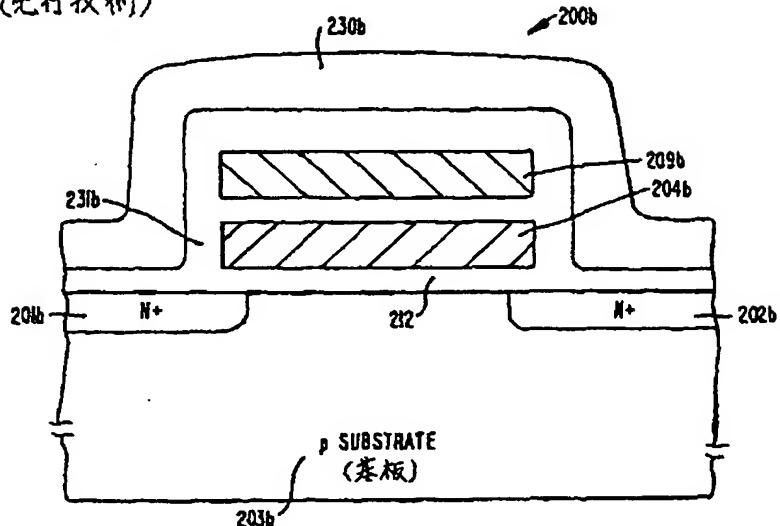
第13b図は表2を含んでおり、この表は第11b図の下層設

置アレイのすべての動作モードの電圧条件を示している。  
 500a……トランジスタ  
 501a……ソース  
 502a……N+ ドレイン  
 503b

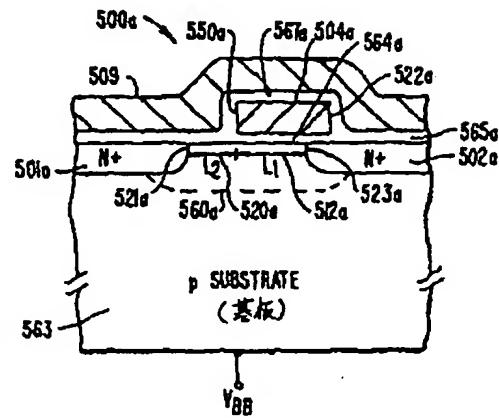
504a……フローティングゲート  
 509a……コントロールゲート  
 560a……チャンネル  
 563a……p 基板

【第1図】

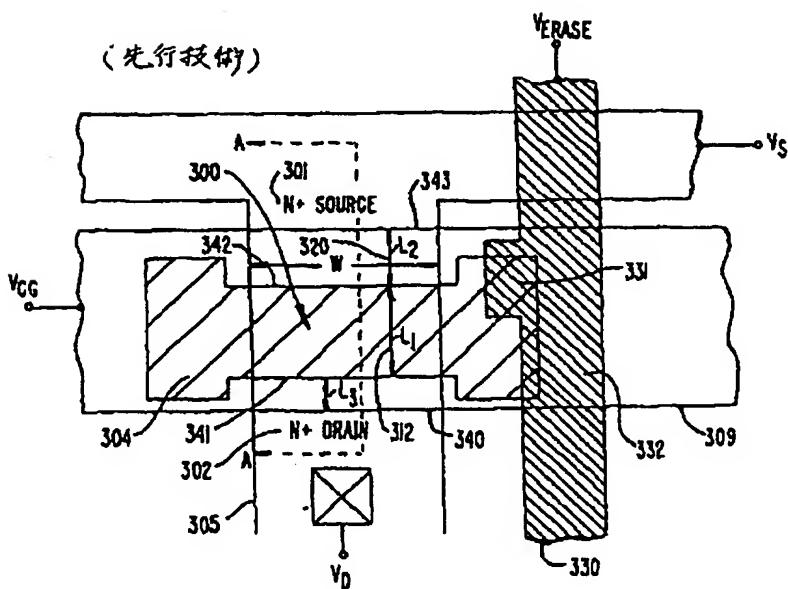
(先行技術)



【第4a図】

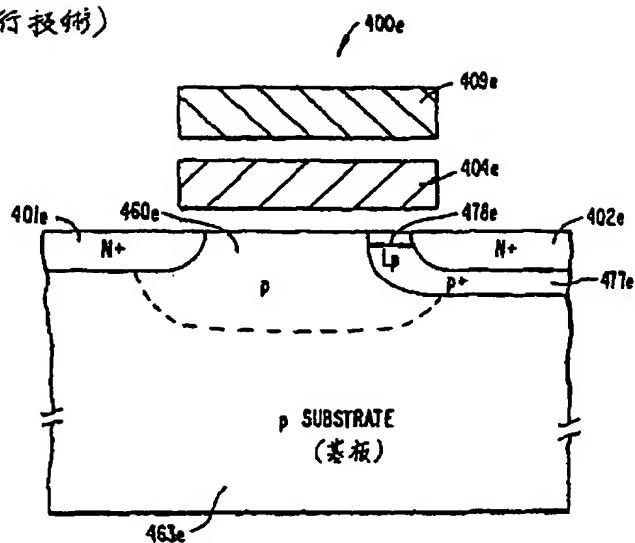


【第2図】

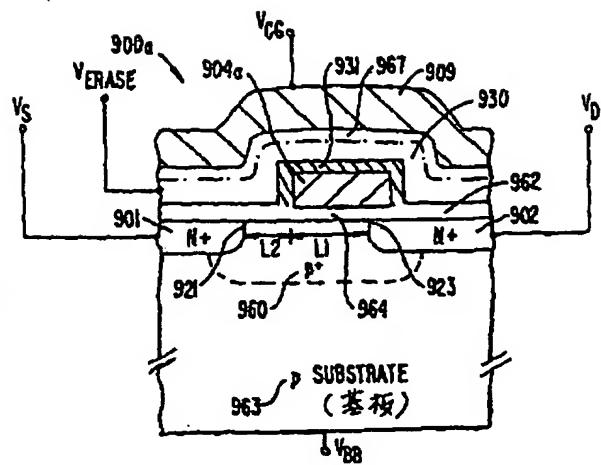


【第3図】

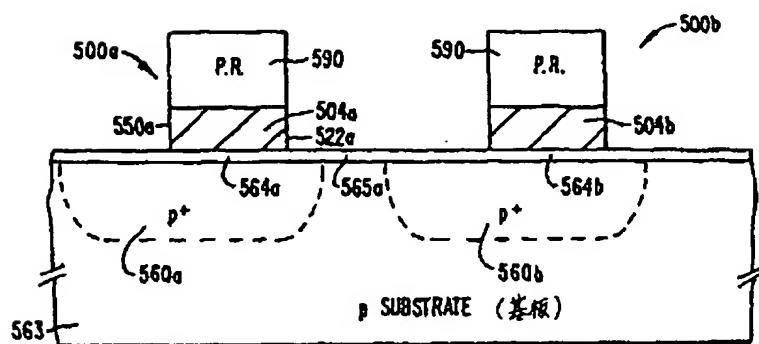
(先行技術)



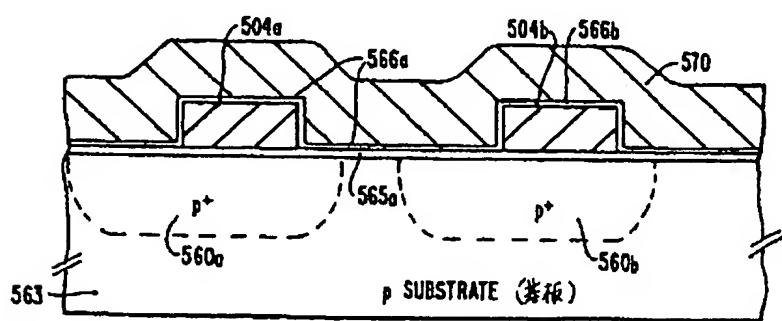
【第8c図】



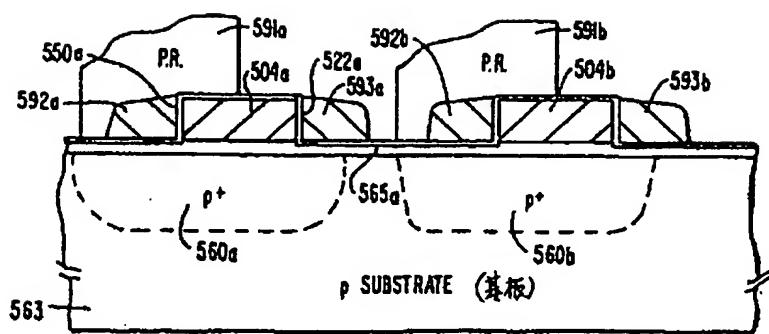
【第4b図】



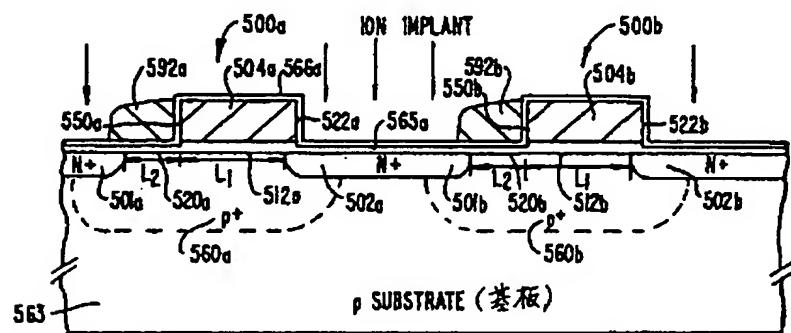
【第4c図】



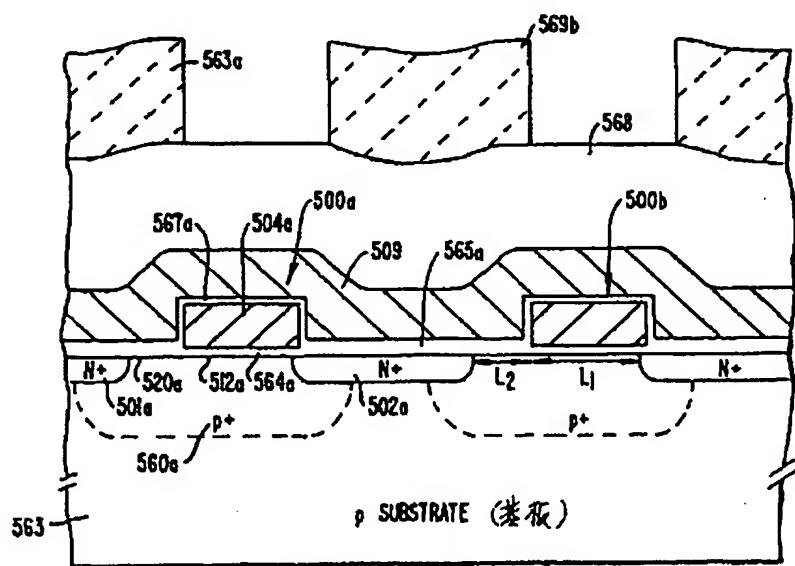
【第4d図】



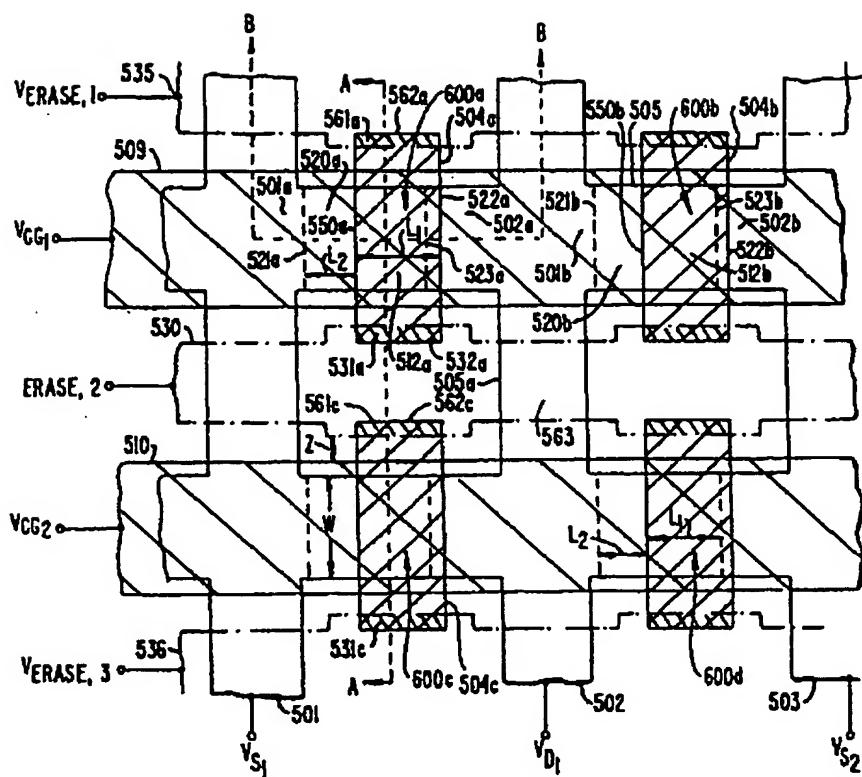
【第4e図】



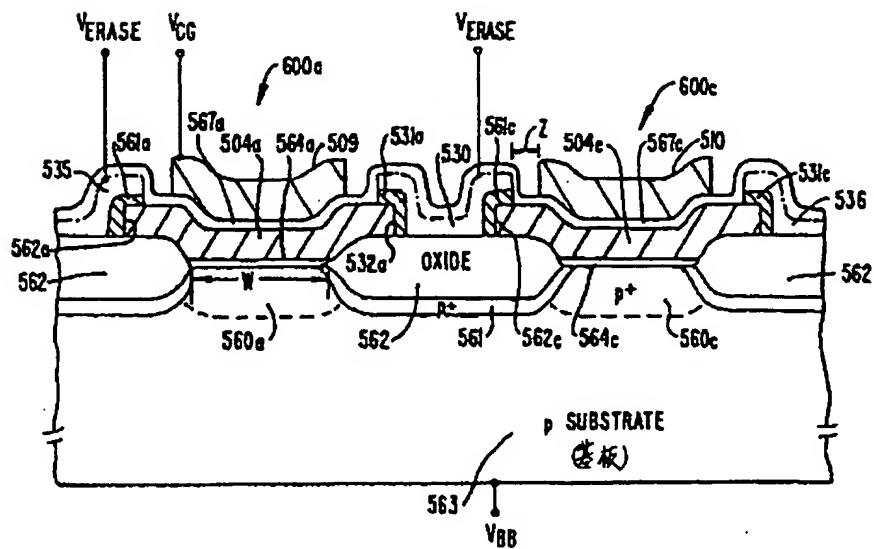
【第4f図】



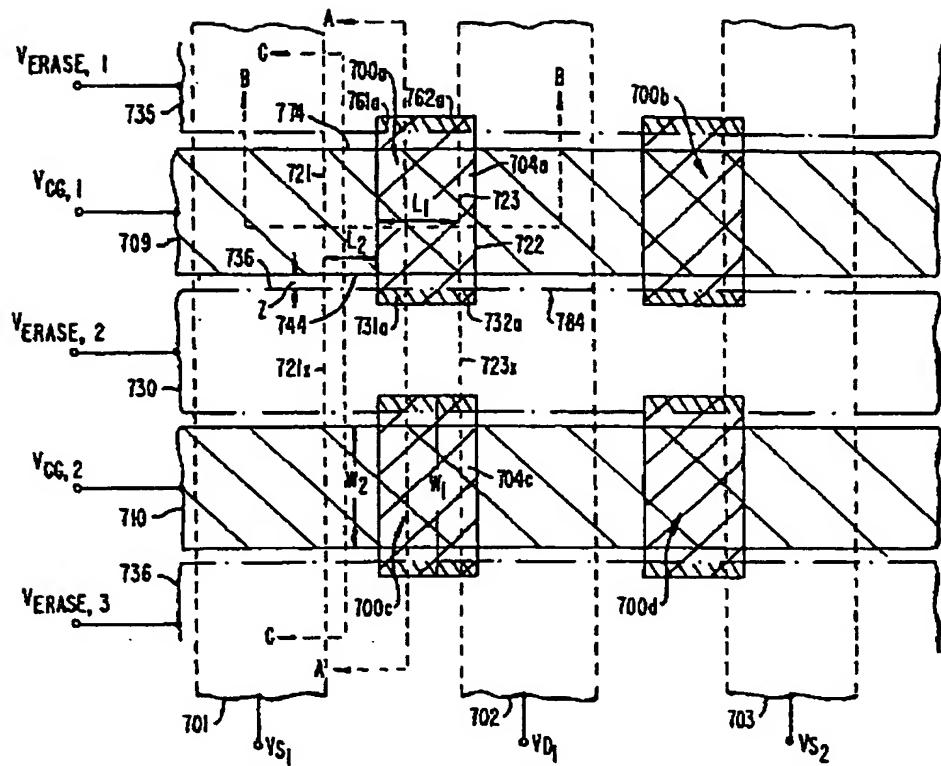
【第5a図】



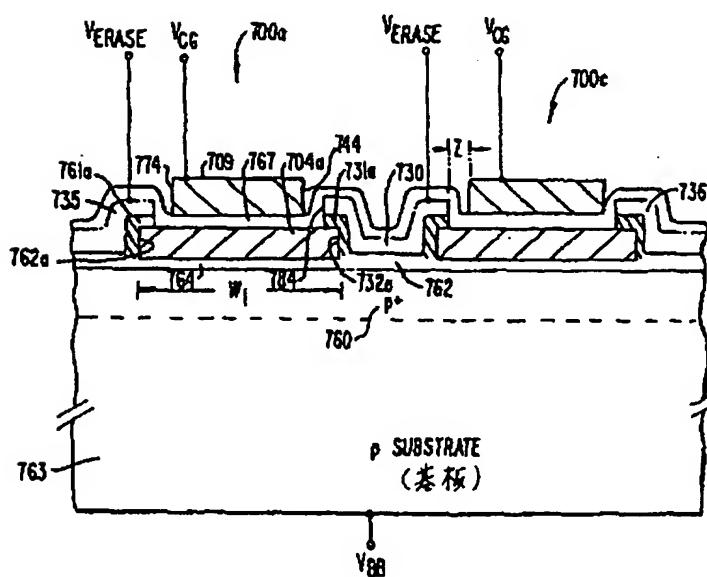
【第5b図】



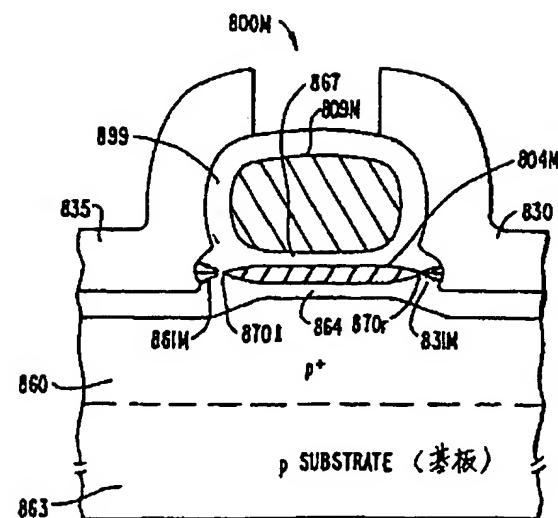
【第6a図】



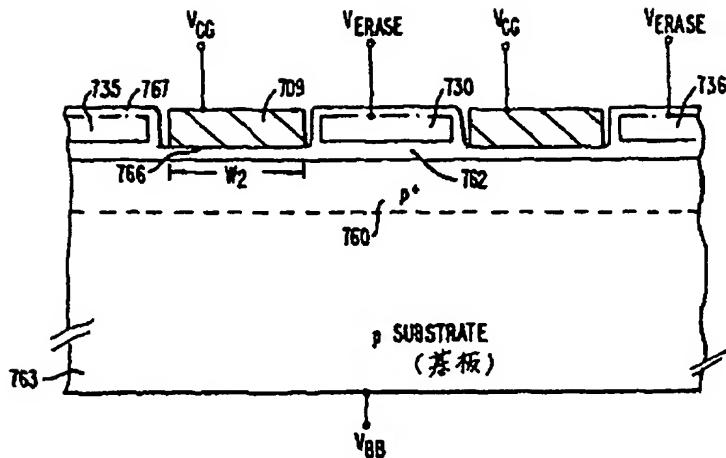
【第6b図】



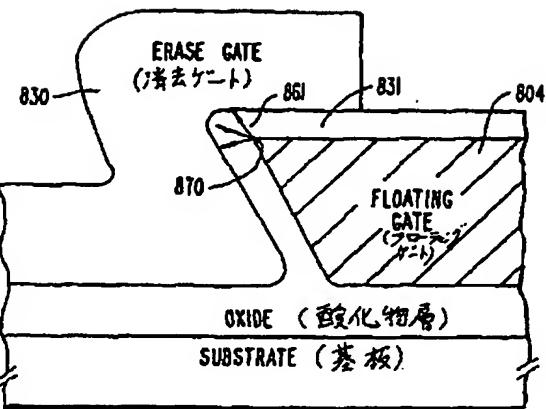
【第12b図】



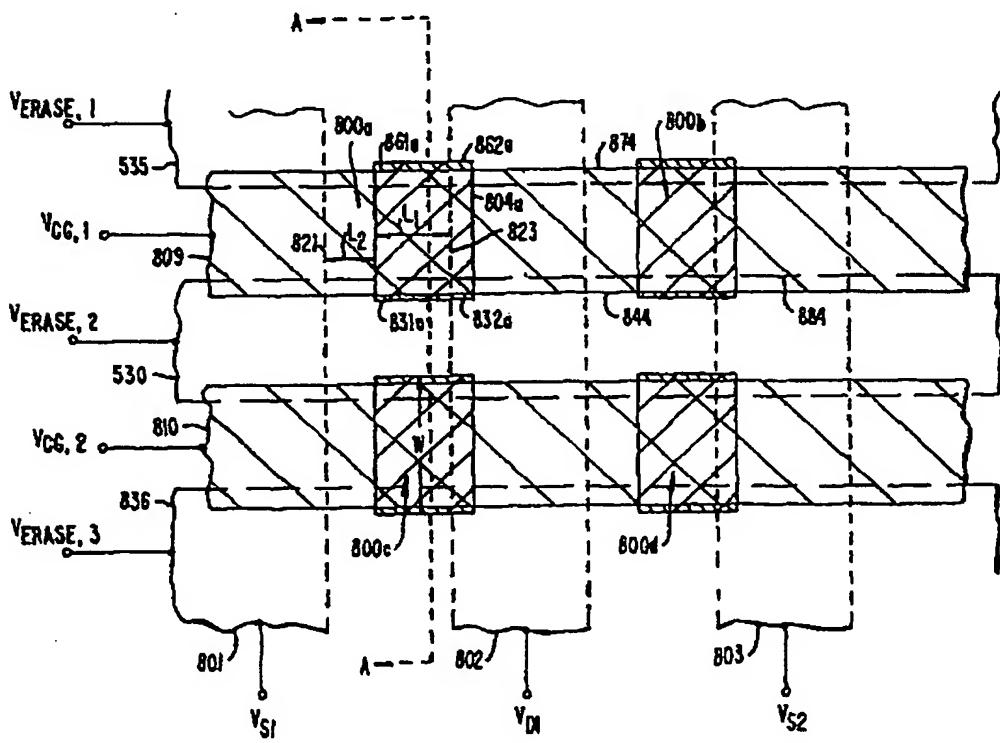
【第6c図】



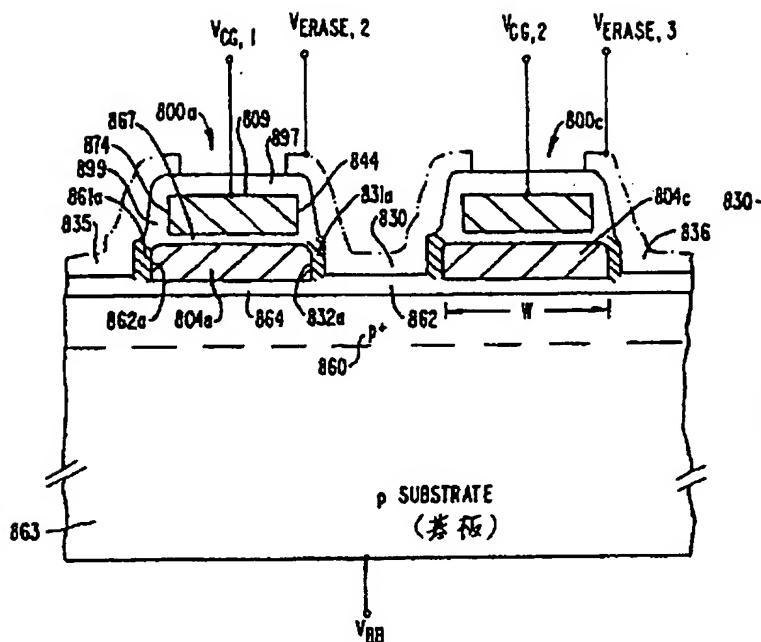
【第12c図】



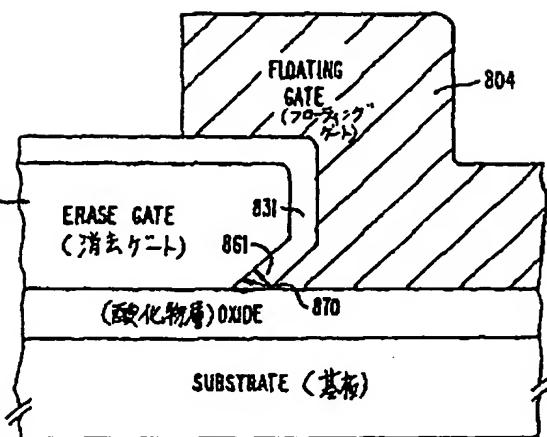
【第7a図】



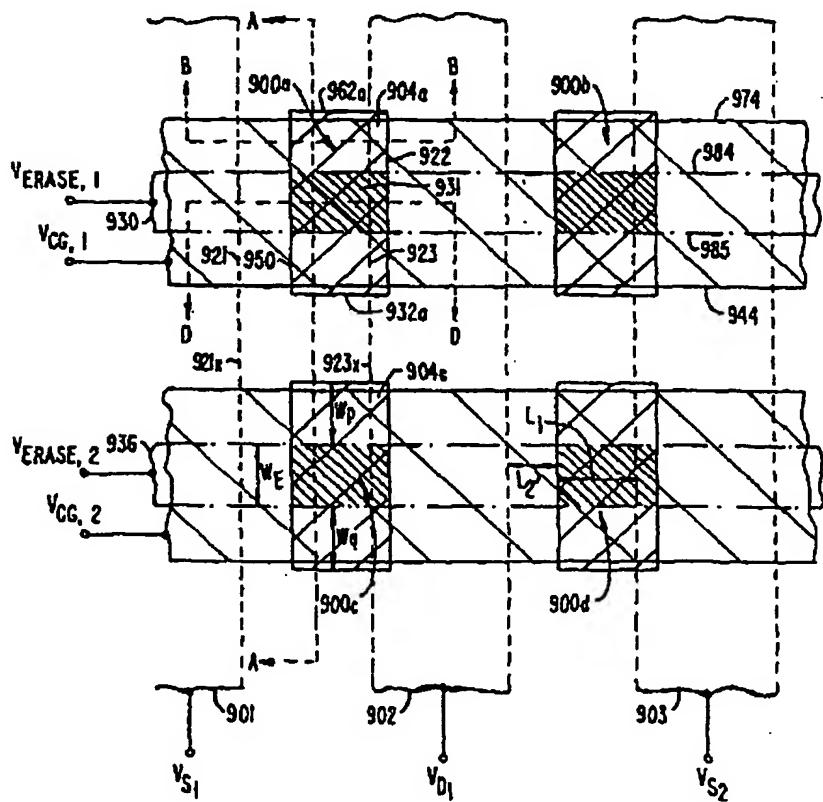
【第 7 b 図】



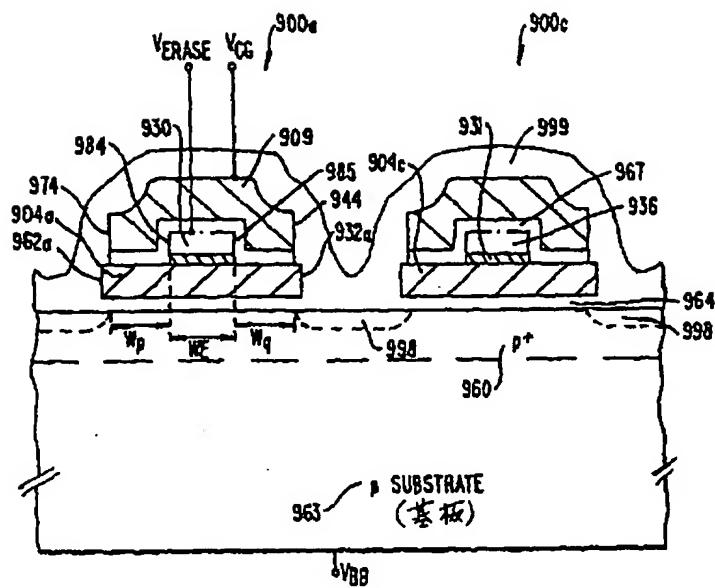
【第 12 d 図】



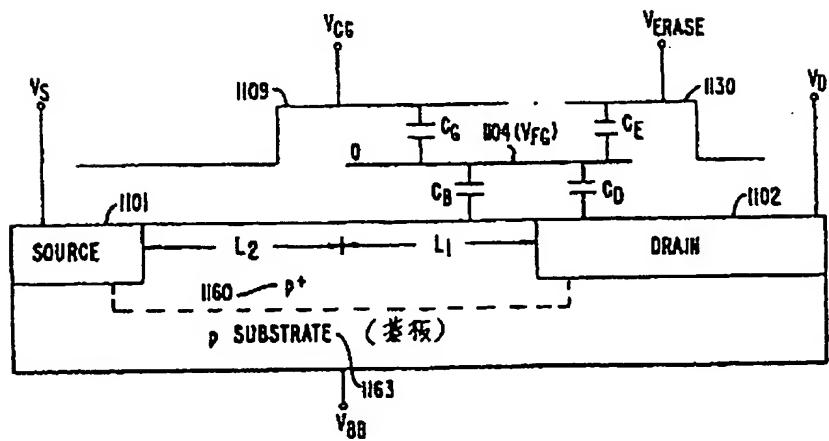
【第 8 a 図】



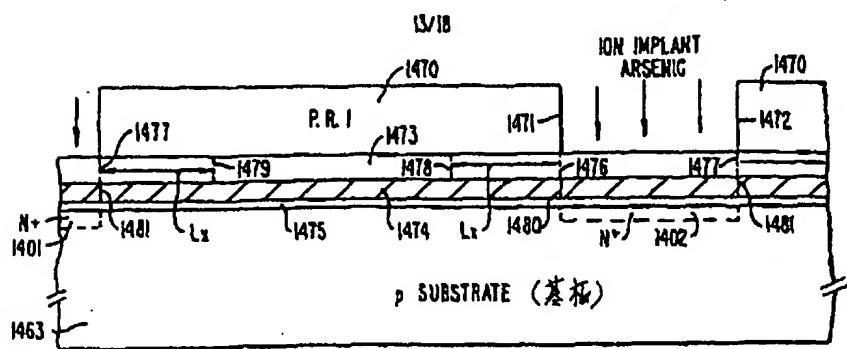
【第8b図】



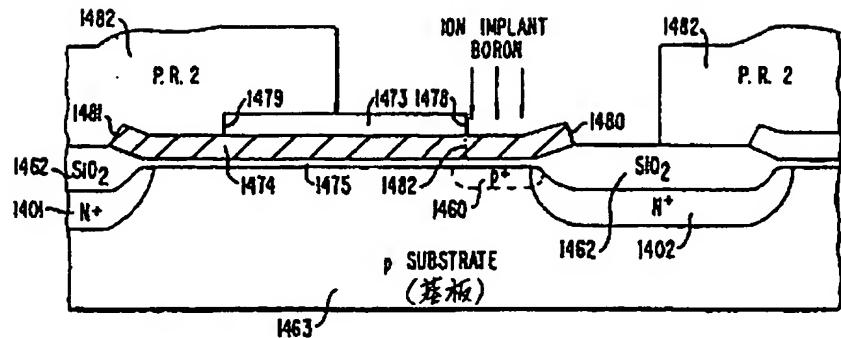
【第9図】



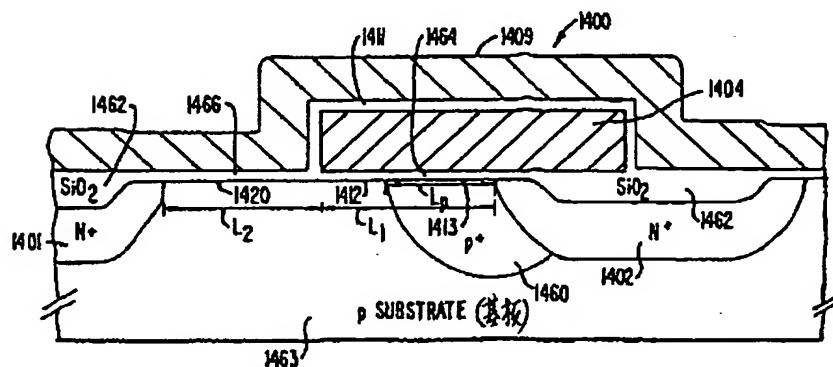
【第10a図】



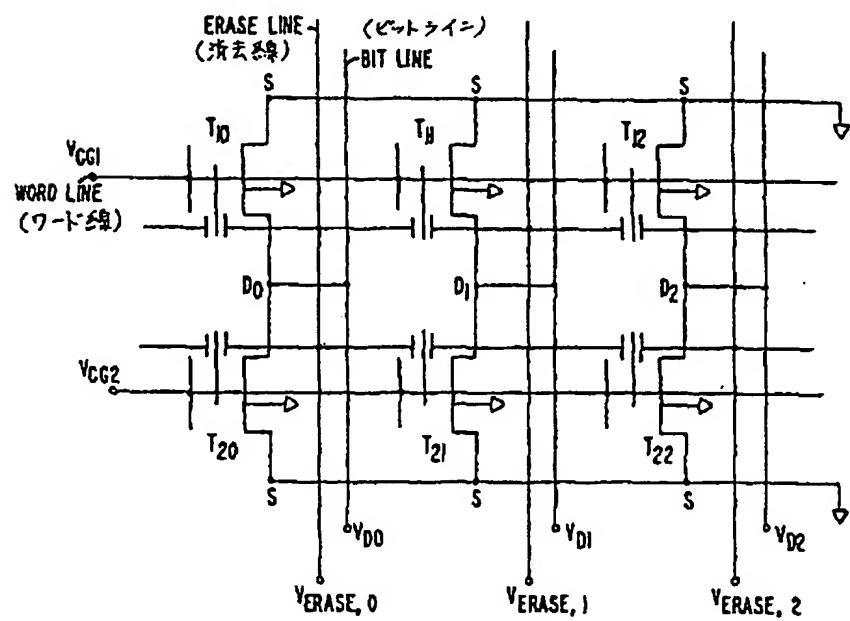
【第10b図】



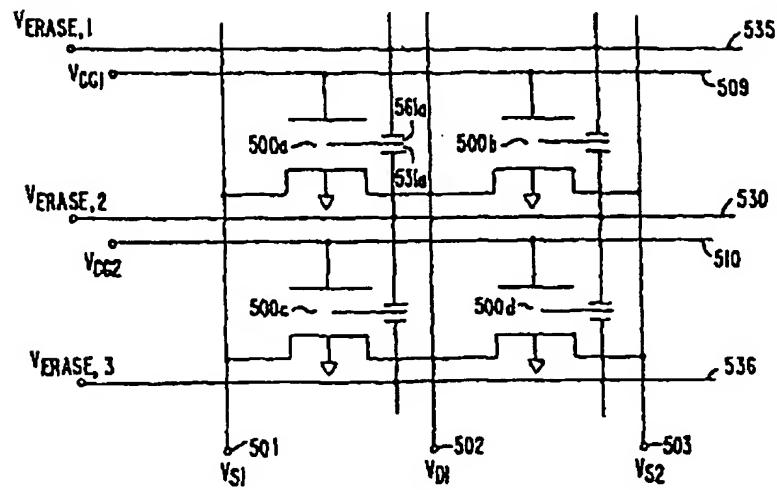
【第10c図】



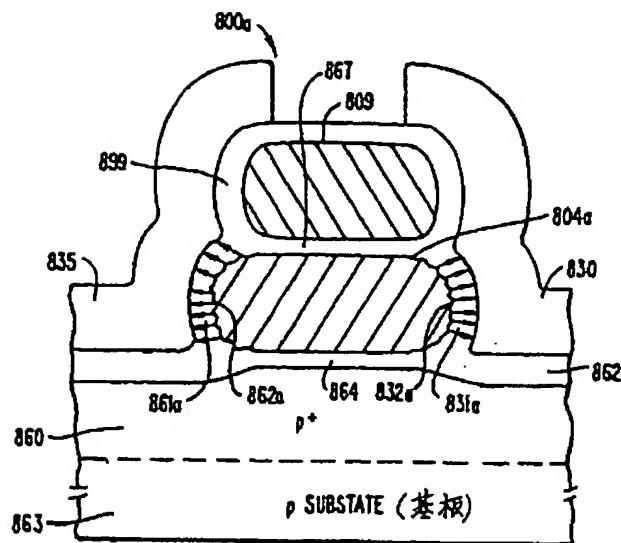
【第11a図】



【第 1 1 b 図】



【第 1 2 a 図】



【第13a図】

TABLE I. ARRAY OF FIGURE IIa. (図11aのアレイ)

	V <sub>CG</sub>			V <sub>D</sub>			S			V <sub>ERASE</sub>		
	SEL. CELL	UNSEL. CELL		SEL. CELL	UNSEL. CELL		SEL. CELL	UNSEL. CELL		SEL. CELL	UNSEL. CELL	
		SAME ROW	SAME COL		SAME ROW	SAME COL		SAME ROW	SAME COL		SAME ROW	SAME COL
READ	5V	5V	0V	1.5V	0V	1.5V	0V	0V	0V	0V	0V	0V
ERASE	0V	0V	0V	0V	0V	0V	0V	0V	0V	20V (BLOCK ERASE) (ブロック 消去)	20V (BLOCK ERASE) (ブロック 消去)	20V (BLOCK ERASE) (ブロック 消去)
PROGRAM	12V	12V	0V	8V	0V	8V	0V	0V	0V	0V	0V	0V

【第13b図】

TABLE II. VIRTUAL GROUND ARRAY OF FIGURE IIb. (図11bの仮想接地アレイ)

	V <sub>CG</sub>			V <sub>D</sub>			S			V <sub>ERASE</sub>		
	SEL. CELL	UNSEL. CELL		SEL. CELL	UNSEL. CELL		SEL. CELL	UNSEL. CELL		SEL. CELL	UNSEL. CELL	
		SAME ROW	SAME COL		SAME ROW	SAME COL		SAME ROW	SAME COL		SAME ROW	SAME COL
READ	5V	5V	0V	1.5V	1.5V	1.5V	0V	1.5V	0V	0V	0V	0V
ERASE	0V	0V	0V	0V	0V	0V	0V	0V	0V	20V (BLOCK ERASE) (ブロック 消去)	20V (BLOCK ERASE) (ブロック 消去)	20V (BLOCK ERASE) (ブロック 消去)
PROGRAM	12V	12V	0V	8V	FLOAT	8V	-0V	FLOAT	-0V	0V	0V	0V